252924-MS-CF

(1)

Congress

1988

Electronic Materials and Processing

Edited by
Prabjit Singh

Published by
ASM

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

REPORT DOCUMENTATION PAGE		READ INSTRUCTIONS BEFORE COMPLETING FORM
1. REPORT NUMBER ARO 25297.14-MS-CF	2. GOVT ACCESSION NO. N/A	3. RECIPIENT'S CATALOG NUMBER N/A
4. TITLE (and Subtitle) 1988 World Materials Congress Proceedings 24-30 Sep 88, Chicago Eight Volumes		5. TYPE OF REPORT & PERIOD COVERED Final 1 Oct 87 - 31 Mar 89 6. PERFORMING ORG. REPORT NUMBER N/A
7. AUTHOR(s) S. G. Fishman and A. K. Dhingra, editors		8. CONTRACT OR GRANT NUMBER(s) DAAL03-87-G-0128
9. PERFORMING ORGANIZATION NAME AND ADDRESS ASM International Detroit, MI 48202		10. PROGRAM ELEMENT, PROJECT, TASK AREA & WORK UNIT NUMBERS N/A
11. CONTROLLING OFFICE NAME AND ADDRESS U. S. Army Research Office P. O. Box 12211 Research Triangle Park, NC 27709		12. REPORT DATE 1988 13. NUMBER OF PAGES
14. MONITORING AGENCY NAME & ADDRESS (if different from Controlling Office)		15. SECURITY CLASS. (of this report) Unclassified 15a. DECLASSIFICATION/DOWNGRADING SCHEDULE

16. DISTRIBUTION STATEMENT (of this Report)

Submitted for announcement only.

17. DISTRIBUTION STATEMENT (of the abstract entered in Block 20, if different from Report)

18. SUPPLEMENTARY NOTES

The view, opinions, and/or findings contained in this report are those of the author(s) and should not be construed as an official Department of the Army position, policy, or decision, unless so designated by other documentation.

19. KEY WORDS (Continue on reverse side if necessary and identify by block number)

Composites, Sheet Steels, Electronic Materials, Wear Resistance, Precipitation Phenomena, High Integrity Castings, Inclusions, HSLA Steels

20. ABSTRACT (Continue on reverse side if necessary and identify by block number)

The proceedings of the 1988 World Materials Congress were published by ASM and consits of the following volumes:

1. Microalloyed HSLA Sttels Conference Proceedings
2. Inclusions and Their Influence on Material Behavior
3. High Integrity Castings
4. Precipitation Phenomena: Deformation and Aging
5. Wear Resistance of Metals and Alloys

(over)

DD FORM 1 JAN 73 1473 EDITION OF 1 NOV 65 IS OBSOLETE

UNCLASSIFIED

SECURITY CLASSIFICATION OF THIS PAGE (When Data Entered)

ABSTRACT CONTINUED:

6. Electronic Materials and Processing,
7. Corrosion-Resistant Automotive Sheet Steels,
8. Cast Reinforced Metal Composites.

Price $ 75.00

Electronic Materials and Processing

Proceedings of the
First Electronic Materials and Processing Congress
held in conjunction with the
1988 World Materials Congress
Chicago, Illinois, USA
24-30 September 1988

Edited by
Prabjit Singh

Sponsored by the
Electronic Materials
and Processing Division of
ASM INTERNATIONAL™

Published by

The publication of this Conference Proceedings of the 1988 World Materials Congress has been made possible by the generous contributions of the U.S. Bureau of Mines, the U.S. Army Research Office, Cytemp Specialty Steel, ELETROMETAL S.A., and Dofasco Inc.

Library of Congress Catalog Card Number: 88-071718
ISBN: 0-87170-341-6
SAN: 204-7586

Printed in the United States of America

ORGANIZING COMMITTEE

Dr. Prabjit Singh
IBM Corporation
Poughkeepsie, NY

Dr. William B. Snyder, Jr.
Oak Ridge National Labs
Oak Ridge, TN

Professor Shyam Murarka
Rensselaer Polytechnic Institute
Troy, NY

Dr. Wei T. Shieh
General Electric Company
Utica, NY

Dr. James R. Lloyd
IBM Corporation
Fishkill, NY

Dr. Jesse H. Chen
Cabot Corporation
Boyertown, PA

Professor Don Preiss
North Carolina State University
Raleigh, NC

FOREWORD

The objective of the First Electronic Materials and Processing Congress, organized by the Electronics Materials and Processing Division (EMPD) of ASM INTERNATIONAL, has been to provide an overview of the field of electronic materials by invited speakers who are leaders in their respective specializations. We have tried to cover as much of the field of electronic materials as is possible in four and a half days of programming. The major subjects covered are semiconductor materials, thin-film dielectrics and metallization, materials for packaging including ceramic substrates and printed wire boards, interconnections and hermetic sealing, connector technology, polymers in electronics and advances in electronic materials.

It is hoped that this Congress will be useful to both the specialists and the generalists practicing in the field and to those who are not in the field but are curious about it and may want to get some tidbits of information they could use in their own fields of specialization.

The Congress was organized by a committee consisting of:

Dr. Prabjit Singh
IBM Corporation
Poughkeepsie, NY

Professor Shyam Murarka
Rensselaer Polytechnic Institute
Troy, NY

Dr. William B. Snyder, Jr.
Oak Ridge National Labs
Oak Ridge, TN

Dr. James R. Lloyd
IBM Corporation
Fishkill, NY

Dr. Wei T. Shieh
General Electric Company
Utica, NY

Professor Don Preiss
North Carolina State University
Raleigh, NC

Dr. Jesse H. Chen
Cabot Corporation
Boyertown, PA

We wish to recognize the contributions of Dr. R. A. Laudise and Dr. G. Y. Chin of AT&T Bell Labs, Murray Hill, NJ; Mr. R. Moore of Perkin Elmer Corporation, Edison, NJ; Mr. C. MacKay of Microelectronics Computer Corporation, Austin, TX; and Dr. C. Sheldon Roberts, a consultant in private practice. We would also like to acknowledge the excellent support provided by the staff of ASM INTERNATIONAL. Without their help, this Congress and its proceedings would not have been possible.

Prabjit Singh
Chairman,
First Electronic
Materials & Processing Congress

TABLE OF CONTENTS

MATERIALS FOR ELECTRONIC PACKAGING

POLYMERS IN ELECTRONICS

ADVANCES IN ELECTRONIC MATERIALS

ARTIFICIALLY STRUCTURED THIN-FILM MATERIALS AND INTERFACES

V. Narayanamurti
Sandia National Laboratories
Albuquerque, New Mexico, USA

ABSTRACT

The ability to artificially structure new materials on an atomic scale by using advanced crystal growth methods such as molecular beam epitaxy and metal-organic chemical vapor deposition has recently led to the observation of unexpected new physical phenomena and to the creation of entirely new classes of devices. In particular, the growth of materials of variable band gap in technologically important semiconductors such as GaAs, InP, and silicon will be reviewed. Recent results of studies of multi-layered structures and interfaces based on the use of advanced characterization techniques such as high-resolution transmission electron microscopy and scanning tunneling microscopy will be presented.

ADVANCES IN SOLID-STATE ELECTRONICS can usually be traced to a combination of new concepts, materials, and device principles. Integrated circuits, for example, were made possible by the discovery of zone refining, which made possible the synthesis of ultra-pure bulk silicon and germanium.

A number of spectacular advances have been made in constructing new materials [1] by ultra-high vacuum and vapor-phase crystal growth techniques. New combinations of materials called heterostructures can be made that have an artificial periodicity and structure which can be controlled at the atomic level. These novel materials include a variety of combinations of metals, semiconductor, and insulators. Quantum mechanical effects can be seen in these thin films and heterostructures, which have dimensions of only a few hundred angstroms. Thus, one can investigate experimentally the one-dimensional quantization of electrons that is familiar to undergraduate physics students. The quantum tunneling of electrons across such thin films can be observed when the thickness of the films is comparable to the extension of electronic wave functions (10 to 100 Å). The control of the electronic and optical properties of materials at the quantum level will have a major impact on new technology, particularly in the area of opto-electronics [2]. Although much work has been done recently in the area of metal films [3], this article will be confined to a discussion of progress in the area of thin-film semiconductor structures grown on technologically important materials such as GaAs, InP, and silicon.

SEMICONDUCTOR HETEROSTRUCTURES AND EPITAXY

A semiconductor film grows epitaxially if the crystallinity and orientation of the deposited layer are determined by the substrate. If the single crystal formed consists of thin films of dissimilar semiconductors one on top of the other, then the process is called heteroepitaxy. Heteroepitaxy provides new degrees of freedom on the surface of a semiconductor. Although the dominant material in semiconductor electronics has been silicon, researchers are striving to improve its electronic properties through controlled changes in the band structure of the material. For example, compound semiconductors such as GaAs have an intrinsically higher electron mobility than silicon. More importantly, unlike silicon, many III-V semiconductors (semiconductors made from elements in groups III and V, e.g., GaAs) have direct energy band gaps that facilitate the efficient recombination of electrons and holes to generate light.

Figure 1 shows a plot of the energy band gap as a function of lattice constant for several III-V semiconductors. The lines that connect points on the graph show how the band gap and the lattice constant vary for mixtures

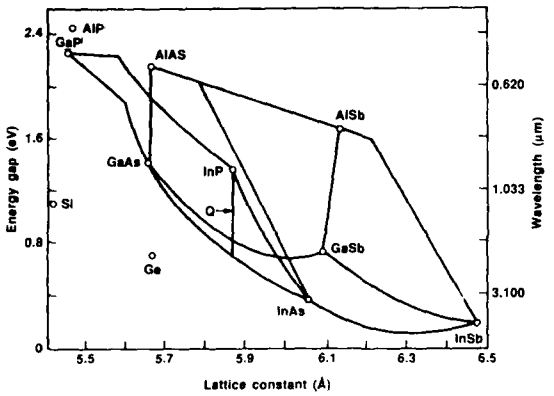


Fig. 1. A plot of the energy gap versus lattice constant for several III-V semiconductors. The lines that connect points on the graph show how the band gap and the lattice constant vary for mixtures of the binary compounds and for quaternary compounds (Q) to which the points correspond. Also shown are the band gap and lattice constant for the elemental semiconductors silicon and germanium.

of the binary compounds to which the points correspond. For example, the ternary compound $Ga_xAl_{1-x}As$, which is closely lattice-matched to GaAs, can be made to have a band gap between 1.4 and 2.2 eV by varying the ratio of gallium to aluminum. The difference in the band gap can be used to confine light and carriers in a $GaAs/Al_xGa_{1-x}As$ heterostructure. Quaternary compounds (Q), such as $In_xGa_{1-x}As_yP_{1-y}$, can be lattice matched to InP and have band gaps in the wavelength range of 1.3 to 1.6 μm, which is the range such that light loss from silica-glass fibers (used in fiber-optic communication) is low. Also shown in Figure 1 are the lattice constants and band gaps of elemental silicon and germanium. Silicon has the smallest lattice constant of the semiconductors shown. Thin-film epitaxy on its surface is only possible through the generation of large compressive stresses that cause the alloy layer to deform and thus match the atomic spacing of the substrate.

FABRICATION OF LAYERED HETEROSTRUCTURES

Great advances in the fabrication of layered heterostructures have taken place in recent years. Among the techniques that have been used are liquid-phase epitaxy (LPE), chemical vapor deposition (CVD), and molecular-beam epitaxy (MBE). In LPE, the epitaxial layer is grown by cooling a heated metallic solution that is saturated with the components needed to grow the layer and that is kept in contact with the substrate. In CVD, the epitaxial layer is grown from a heated source of gaseous elements or compounds that react at the substrate surface. Considerable progress has been made in the growth of quantum well heterostructures with metal-organic chemical vapor deposition (MOCVD). Metal alkyls are used as the compound source. In MBE, the deposition occurs under controlled ultra-high vacuum conditions. Crystal growth results through the reaction of one or more thermal beams of atoms and molecules of the constituent elements with a crystalline substrate held at a suitable temperature. The growth is monitored through *in situ* surface diagnostic tools, such as low-energy electron diffraction (LEED). The clean environment, the slow growth rate, and the independent control of the beam sources allow the precise fabrication of atomic scale semiconductor heterostructures.

The clear distinctions among the growth techniques have recently begun to fade, and new techniques have appeared. In the case of III-V compound semiconductors, the high vapor pressure of the group V sources, particularly phosphorus, has led to the development of gas-source MBE (GSMBE). In this hybrid technique [4], the gases AsH_3 and PH_3 are passed through a thermal, catalytic cracker oven and then introduced into the vacuum chamber through a controlled leak. In this way, a stable beam intensity is obtained. For growth of InGaAsP, precise control of the mix of arsenic and phosphorus is possible. In a further variant [5] of GSMBE, the group III elements are organometallic compounds as in MOCVD. Hybrid methods that use the advantages of an ultra-high vacuum environment and of gaseous sources will likely become the techniques of choice.

Progress in these crystal growth techniques has allowed the construction of atomic scale structures that can have virtually arbitrary potentials for electrons and holes. These potentials include abrupt band discontinuities, which are used to confine carriers in a two-dimensional (2-D) state. Because of their applications to novel and potentially useful devices, the physics of thin-film semiconductor heterostructures has become a subject of intense study.

HIGH CARRIER MOBILITY STRUCTURES

The principal building block of silicon integrated circuits (ICs) is the metal-oxide field-effect transistor (MOSFET), which can be described simply as a layer of 2-D electrons (holes) confined to the interface between crystalline silicon and its native oxide (SiO_2). In the case of GaAs, the inversion layer can be formed at the interface of a $GaAs/Al_xGa_{1-x}As$ heterostructure. Until 1978, carrier mobilities in such structures were extremely low. At that time, Stormer and his colleagues introduced modulation doping [6], in which the band-edge discontinuity in the heterostructure is used to separate the electrons (holes) from their parent donors (acceptors), as shown in Figure 2A. This results in greatly enhanced mobilities (μ) when compared with uniformly doped materials. The greatest improvement in the mobility is found at low temperatures, where scattering from ionized impurities is the dominant mechanism for reducing mobility in uniformly doped materials. Even at room temperature,

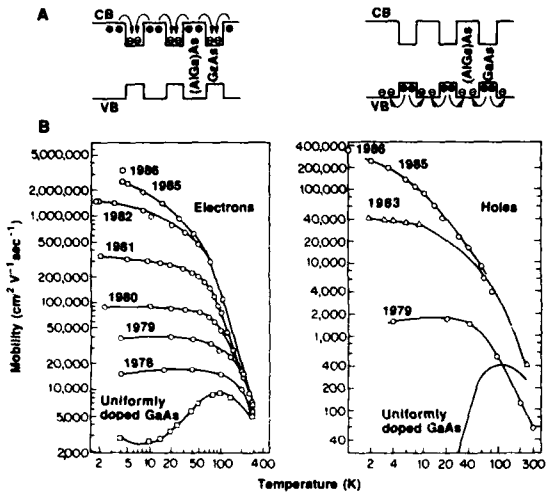


Fig. 2. (**A**) Schematic illustration of modulation doping for electrons and holes. (The conduction and valence bands are indicated as (CB and VB.) (**B**) Highest reported electron and hole mobilities as function of temperature during the period 1978 to 1986 (compiled by A. C. Gossard).

modulation-doped structures show an enhancement in mobilities over a uniformly doped material, in spite of the fact that at higher temperatures thermal phonon scattering becomes progressively more important.

Figure 2B shows a plot of the electron and hole mobility that could be achieved with modulation-doped structures during the period 1978 to 1986. The steady improvement in the mobility that could be achieved can be traced to: the introduction of an undoped spacer layer adjacent to the barrier layer that further separates the parent donors (acceptors) from the electrons (holes); the use of a modulation-doped single interface instead of a superlattice; and improvements in growth techniques, such as the introduction of a sample load lock to prevent contamination of the main growth chamber during introduction of the sample. The achievement of high carrier mobilities has led to the observation of novel quantization phenomena (the fractional quantum Hall effect) for 2-D electron and hole systems [7].

Modulation doping causes greatly enhanced conductivity in the plane of the layered structure as compared to that in the direction perpendicular to the layers. This effect is exploited in the modulation-doped field-effect transistor (MODFET). These devices are characterized by high switching speeds and low power consumption. At 77 K, where the modulation doping advantage is already significant, a switching time as short as 5.8 ps has been observed [8]. Such transistors will probably form the basis of advanced GaAs digital circuits. A 4-kilobit static random access memory (RAM) and a high-speed multiplier have been fabricated. The recent demonstration of high-mobility holes has led to the fabrication of low power, complementary (n- and p-type) transistor circuits in the GaAs system.

QUANTUM STATES, PERPENDICULAR TRANSPORT, AND TUNNELING

Quantum wells in GaAs with thicknesses of a few hundred angstroms can result in the quantum confinement of electrons. These quantum states were observed in 1974 at low temperatures in electron transport through resonant tunneling [9] and through optical absorption [10] at well-defined energies. The quantum confinement also causes modification of the electronic density of states and has led to the observation of several novel effects, such as room temperature exciton absorption and low-threshold semiconductor lasers.

An example of a structure which illustrates the effects of perpendicular resonant tunneling is shown in the upper part of Figure 3, which is a schematic of the first room temperature [11] resonant tunneling bipolar transistor (RTBT). The wide band-gap $Al_xGa_{1-x}As$ emitter is graded [12] by chopping the beam from the aluminum effusion cell over a distance of about 500 Å to a value of x = 0.07, so that the energy difference between the emitter electrons and the first quantum state in the base of the transistor is near the thermal energy of the electrons. The base itself consists of a 74 Å, undoped GaAs well with 21.5 Å AlAs barriers. Electrons are thermally injected into and transported through the base. The collector current (I_c) (at a fixed collector-emitter voltage) increases with a gain of ~5 as a function of base current (I_b) for $I_b < 2.5$ mA, which is a normal characteristic of a bipolar transistor.

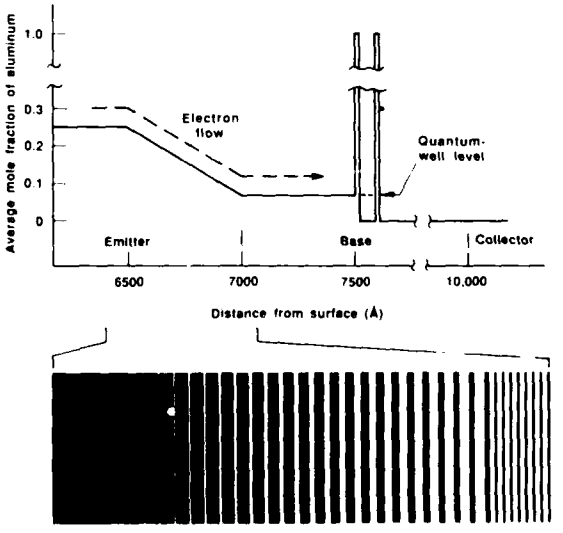


Fig. 3. Resonant tunneling bipolar transistor (RTBT) in the GaAlAs system grown by MBE. The aluminum composition of the emitter is graded by chopping of the effusion cell so that the average composition changes smoothly from a value of 0.25 to 0.07 and so that the emitter electrons are in resonance with the first quantum level in the base (*11*).

At higher base currents, the current gain is quenched approximately exponentially as a result of the suppression of resonant tunneling, because the conduction band edge of the $Al_{0.07}Ga_{0.93}As$ is above the first energy level of the quantum well. In the region of negative transconductance, the RTBT behaves as an oscillator. The number of possible [13] transistors with unique functions is large. Through careful design of structures with several resonant states in the base, it should be possible to fabricate bipolar transistors with multiple stable states.

GAS-SOURCE MBE AND QUANTUM WELLS IN InGaAsP

There has been significant progress in the growth of heterostructures and quantum wells on InP with the use of GSMBE and low-pressure MOCVD. The entire Q range of quaternary compositions (see Figure 1) of $Ga_xIn_{1-x}As_yP_{1-y}$ on InP of interest for 1.3 to 1.6 μm fiber-optic transmission and heterostructure electronic devices, such as optical modulators and detectors, have been grown.

The degree of control [14] that has been achieved in crystal growth is revealed with the transmission electron microscope (TEM) image shown in Figure 4. Micrographs of four $Ga_{0.47}In_{0.53}As$ quantum wells with thicknesses of 6, 9, 12, and 24 Å separated by InP barriers (which are ~150 Å thick) show that there are fluctuations of about one monolayer at the layer interfaces. Similar results were obtained with the quaternary quantum wells, although in that case, the interfaces seem even more abrupt. Figure 4 shows measurements of low-temperature photo-luminescence of these single quantum wells after illumination with light of higher energy than the band gap (632 nm, or 1.95 eV). The figure illustrates the enormous quantum size effect in the quantum well emission compared to that of bulk InGaAs (~1.6 μm) which for the thinnest wells (6 Å) corresponds to a shift in energy ΔE of 550 meV. For the 6 Å well, the wavelength of emission is about 0.94 μm. Thus, by choosing wells of different thicknesses, one can make quantum well laser structures that span a wide wavelength range. The optical data also suggests a high degree of structural perfection (monolayer abruptness).

LATERAL QUANTUM CONFINEMENT

MBE has been primarily used to provide quantum-confined structures in the direction of growth. Confinement in one and zero dimensions (quantum wires and quantum boxes) requires fabrication of laterally confined structures on a scale of about 100 Å. Such laterally confined structures can, in principle, be made by lithographically patterning the transverse dimensions of a quantum well structure grown by MBE. Such dimensions are at the present limits of

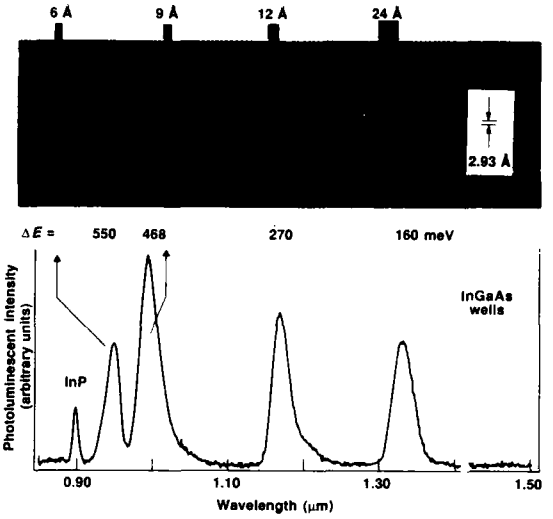


Fig. 4. Transmission electron microscope image of InGaAs quantum wells of thickness 6, 9, 12, and 24 Å separated by 150-Å InP barriers. The bottom shows photoluminescence spectra taken at 6 K of the quantum wells. The shift ΔE is measured from the known energy gap of bulk InGaAs (*14*).

electron-beam lithography. In small structures, the ratio of surface to volume is large and the patterning must be such that it does not degrade the sidewalls of the layers. Because of modification of the electronic density of states in these confined structures, several novel optical effects, such as an improvement in the performance of quantum wire semiconductor lasers, have been theoretically predicted.

Progress has been made in fabricating such structures in both the GaAs/AlGaAs and the InGaAs/InP systems. Gibert, *et al.* [15], have fabricated quantum wires and boxes. They used gallium ions to disorder local areas of a GaAs/AlGaAs quantum well structure (see Figure 5). The ion implantation occurred through open areas of an electron-beam defined tungsten mask. They discovered that the inter-diffusion of aluminum in GaAs after a rapid thermal anneal is much larger in the disordered region than in the masked region. Thus, they were able to produce high-quality, laterally confined structures, in which new optical transitions that correspond to the theoretically predicted energy levels for quantum wires and boxes were observed.

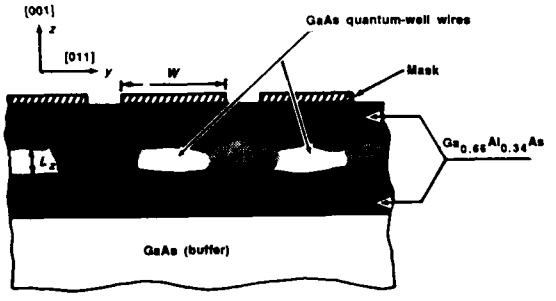


Fig. 5. Schematic (not to scale) of GaAs quantum well wire structure after implantation and annealing. The position of the mask (which has a width W) during implantation is indicated. The wires (which have a thickness L_z) are actually 1 μm apart (*15*).

An interesting case for the study of lateral confinement is the InGaAs/InP system. In this system, the surface recombination velocity of electrons and holes is ~0.01 of that in the GaAs/AlGaAs system [16]. Thus, high luminescence efficiency may be expected. Temkin, *et al.* [17], have fabricated quantum wires and boxes on the order of 300 Å in this system with electron-beam lithography. Figure 6A shows an electron micrograph of some typical quantum boxes with an average diameter of 300 (±50) Å. Quantum wires with lateral dimensions ~300 Å have also been fabricated. Figure 6B shows typical photo-luminescence spectra for these quantum structures. The luminescence shift of 8 to 14 meV is consistent with that expected for lateral dimensional confinement. The high photo-luminescence efficiency is also consistent with the low surface recombination velocity that is characteristic of InP and its alloys.

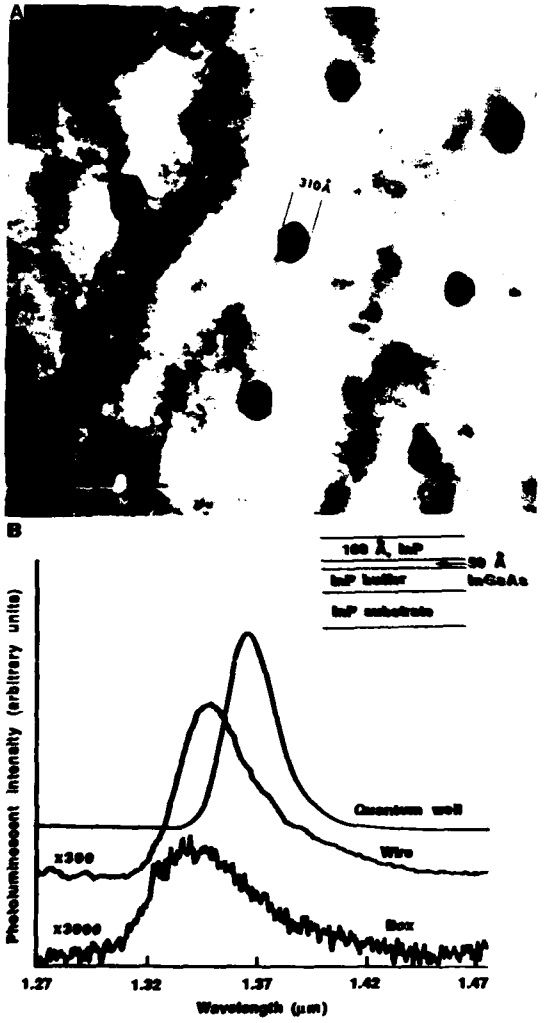


Fig. 6. (A) The transmission electron micrograph shows InGaAs/InP quantum boxes that were fabricated by direct electron-beam writing and ion-beam milling. The average box diameter is 300 Å. The initial 50-Å InGaAs quantum well, upper panel, was grown by GSMBE (*17*). (B) Photoluminescence of control quantum well sample, and for lithographically patterned 300-Å quantum well, wires, and boxes. The wires and boxes have thicknesses of ~300 Å.

HETEROEPITAXY ON SILICON

The use of silicon in semiconductor applications is ultimately limited by its incompatibility with other semiconductor materials. Silicon as a substrate material alone has attractive properties such as high mechanical strength, high degree of crystalline perfection, large diameters (~20 cm) to which it can be grown, natural abundance, and low cost. It is clear from Figure 1 that if the electronic and optical properties of the silicon surface are to be altered through lattice-matched heterostructures, the number of possibilities are few. Both GaP and AlP, which are lattice-matched to silicon, are composed of elements that are commonly used as dopants in silicon. Thus, when crystal growth of thin films has been attempted, uncontrolled cross-doping effects (e.g., inter-diffusion of gallium or aluminum into the silicon) have resulted. Two major approaches have been used in heterostructure band-gap engineering on silicon. In the first, the elemental semiconductor germanium and its alloys Ge_xSi_{1-x} are grown. The idea is to use Si/Ge_xSi_{1-x} heterostructure single interfaces and superlattices for altering the band gap and, hence, the electronic and optical properties on the surface of silicon. In the second, silicon is used as a substrate material for growth of $GaAs/Al_xGa_{1-x}As$ heterostructure devices. This approach combines the excellent properties of silicon as a substrate material with the unique electronic and optical properties of GaAs/AlGaAs heterostructures.

Although germanium has the same crystal structure as silicon, its lattice constant is approximately 4% larger. The epitaxial growth can occur either as incommensurate or as commensurate epitaxy. In the former (upper inset of Figure 7), both crystals retain their individual structure; at the interface of germanium and silicon, every 25th row of atoms will have only three bonding neighbors. Rows of these improperly bonded atoms form "misfit dislocations" that can seriously degrade the electronic properties of the interface. In commensurate epitaxy (bottom inset of Figure 7), the fourfold bonding at the interface is preserved by compression of the larger germanium lattice along the interface and by elongation perpendicular to the interface. Such strained-layer epitaxy is possible if the stored strain energy in the film is lower than the reduction in the dangling bond energy.

Heterostructures of Si/SiGe have been grown by MBE by Bean [18]. Figure 7 shows a plot of the thickness of Ge_xSi_{1-x} alloys on silicon for commensurate epitaxy. The observed thicknesses for commensurate epitaxy are about ten times as large as those predicted by equilibrium theories of dislocation formation. The band gap of the

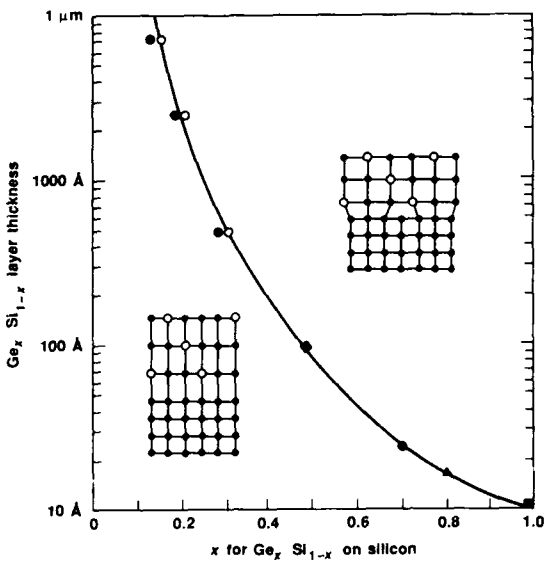


Fig. 7. Plot of critical film thickness for commensurate epitaxy as a function of x for Ge_xSi_{1-x} alloys grown on silicon substrates. The insets show the atomic arrangements for the strained but defect-free material found in sufficiently thin layers and for the bulk material in which misfit dislocations accommodate the lattice mismatch. [Adapted from (*18*)]

strained GeSi films is considerably smaller than that of corresponding bulk alloys. The growth of GeSi heterostructures on silicon thus allows one to exploit "band-gap engineering" techniques for the production of novel, long wavelength photo-detectors and MODFETs with silicon-based technologies.

The other approach is heteroepitaxial semiconductor (GaAs) growth on silicon [19]. Like germanium, GaAs also has a lattice constant some 4% larger than that of silicon. GaAs is a polar semiconductor with a zinc blend structure. Thus, in addition to misfit dislocations, one must contend with control of additional structural defects such as anti-phase domains. An anti-phase domain is a region of the crystal where the nearest-neighbor bond is not between a gallium atom and an arsenic atom, but is between two gallium atoms. The formation of such anti-phase domains can be minimized by using misoriented silicon substrates. A silicon surface that is cut a few degrees off the <100> direction has a predominant number of double atomic steps (steps whose height is twice that of a single atomic step). Growth on such surfaces leads to a material that is free of anti-phase domains.

In order to grow thick layers of GaAs on silicon with a low dislocation density, researchers have tried several other approaches besides misoriented substrates. In one approach [20], a germanium film (which has approximately the same lattice constant as GaAs, see Figure 1) is grown by MBE. A GaAs/GaAsP strained-layer superlattice (SLS) is then grown by MOCVD, over which a thick GaAs film is grown (see Figure 8). It is clear from the transmission electron microscope (TEM) image in Figure 8 that the SLS appears to have trapped dislocations that may have moved from the germanium layer during the growth of the GaAs layer. The apparent dislocation density of these films at the surface is low (~10^4 to ~10^5 cm^{-2}). Such structures have been used to grow GaAs/AlGaAs double heterostructure lasers that have low current thresholds. These lasers are not yet as good as those grown directly on GaAs substrates, but the progress in the growth of such structures, which are extremely sensitive to dislocation, has been remarkable.

MBE also allows the growth of compatible metals, such as silicides and insulators, on silicon. A buried heterostructure [21] of $Si/CoSi_2Si$ has been grown that demonstrates transistor action with the metal laser as a base. Epitaxial fluorides [22] on silicon have been used both for dielectric isolation and for the fabrication of field-effect transistors (FETs). Most recently, ordered monolayer structures [23] of germanium on silicon have been grown, which represent another class of new materials on silicon.

The possibilities opened by heteroepitaxy with MBE are many and have only just begun to be explored. The low growth temperature and control permitted by MBE are unparalleled and will permit much further development in the area of heteroepitaxy.

Fig. 8. Transmission electron micrograph of a GaAs/GaAs-GaAsP SLS structure grown on Ge/Si composite substrate. The GaAs layer is 2 μm thick. The germanium film on silicon was grown by MBE, the GaAs and GaAsP were grown by MOCVD. [Adapted from (*20*)]

CHARACTERIZATION OF INTERFACES AND THIN FILMS

The surface and interface characterization of MBE-grown films has been of major importance in understanding the early stages of crystal growth and lattice formation. The earliest advances were made with *in situ* LEED. More recently, powerful analytical techniques such as high-resolution TEM (HRTEM), x-ray scattering, and Rutherford backscattering have been used to characterize the degree of perfection of thin-film heterostructures. Often the results of a variety of techniques must be combined to obtain a complete picture.

I have highlighted electron microscopy, which gives excellent visual images of the degree of perfection of thin-film interfaces and their structure.

Another microscopic technique that gives detailed and direct information of interface structure is the recently developed scanning tunneling microscope (STM) of Binnig, *et al.* [24]. The STM measures the very small current that flows when a potential is applied between the surface and a probe tip that is scanned across the interface at a distance of only a few angstroms. The quantum mechanical tunneling current is extremely sensitive to the distance between atoms on the surface and the tip, and serves as a direct method of observing individual atoms at interfaces. The STM has recently been used by Golovchenko and co-workers [25] to obtain the first images of atomic steps and their structure on silicon surfaces. They have also imaged the strain-stabilized crystal structure of germanium and GeSi grown by MBE on silicon surfaces.

LEED studies [26] showed that germanium epitaxially grown on silicon has a wide variety of reconstructions that depend on heat treatment. The STM images [27] shown in Figure 9 provide the first detailed atomic scale images of the Ge c(2 x 8), GeSi(5 x 5), and Ge(7 x 7) reconstructions in silicon. (These are lattice symmetry designations.) These results show the remarkable detail of the surface structure and have helped to resolve many outstanding questions on the relation between clean germanium and silicon surface reconstructions. Of particular interest is the observation of the GeSi(5 x 5) reconstruction that has a rhombohedral unit cell with a lattice constant of 19.2 Å. The asymmetry between the two halves of the unit cell is clearly visible and suggests that an ordered alloy with alternate positions around the deep depressions exists on the surface.

These studies, like many of the other cited in this article, are still in their infancy. It is probable that future MBE machines will have capability for *in situ* studies of crystal growth by the most powerful characterization techniques, such as STM and HRTEM. If recent history is any guide, it is likely that such studies will lead to entirely new classes of materials as a fuller understanding is developed of the first stages of crystal growth.

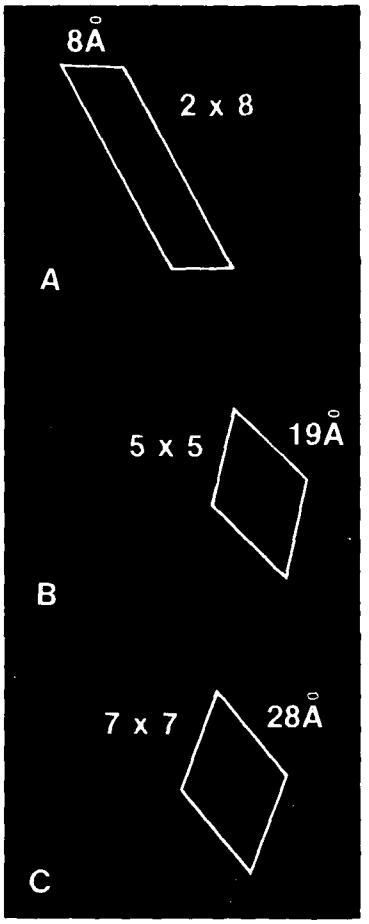


Fig. 9. Scanning tunneling microscopy images of germanium and GeSi on silicon grown by MBE and with different heat treatments. (**A**) The c(2 × 8) reconstruction for germanium on silicon, which is the normal reconstruction for clean germanium. (**B**) The newly observed (5 × 5) reconstruction for a 50:50 alloy of GeSi. This image also shows indications of ordering. (**C**) The (7 × 7) reconstruction, which is typical of clean silicon. [Adapted from (27)]

REFERENCES

1. Cho, A. Y., Thin Solid Films 100, 291 (1983); and Dupuis, R. D., Science 226, 623 (1984).
2. Narayanamurti, V., Phys. Today 37, No. 10, 24 (1984).
3. Falco, C. M., Festkorperprobleme, Advances in Solid State Phys. 25, 531 (1985).
4. Panish, M. B., J. Electrochem. Soc. 127, 2729 (1980); and Panish, M. B., in Proc. of 1986 York Molecular Beam Epitaxy Conf., to be published in J. Cryst. Growth.
5. Tsang, W. T., Appl. Phys. Lett. 45, 1234 (1984).
6. Stormer, H. L., Surf. Sci. 132, 519 (1983).
7. Stormer, H. L., and Tsui, D. C., Science 220, 1241 (1983).
8. Shah, N. J.; Pei, S. S.; Tu, C. W.; and Tiberio, R. C., IEEE Trans. Electron Devices ED-33, 543 (1986).
9. Chang, L. L. and Esaki, L., "Molecular Beam Expitaxy," Pergamon Press, Oxford (1980).
10. Dingle, R., Festkorperprobleme 15, 21 (1975).
11. Capasso, F.; Sen, S.; Gossard, A. C.; Hutchinson, A. L.; and English, J. H., IEEE Electron Device Lett. EDL-7, 573 (1986).
12. Gossard, A. C., IEEE J. Quantum Electron. 22, 1649 (1986).
13. Capasso, F., Science 235, 172 (1987).
14. Panish, M. B.; Temkin, H.; Hamm, R. A.; and Chu, S. N. G., Appl. Phys. Lett. 49, 164 (1986).
15. Cibert, J.; et al., Appl. Phys. Lett. 49, 1275 (1986).
16. Casey, H. C., and Buehler, E., Appl. Phys. Lett. 30, 247 (1977).
17. Temkin, H.; Dolan, G. J.; Panish, M. B.; and Chu, S. N. G., Appl. Phys. Lett., in press.
18. Bean. J. C., Science 230, 127 (1985).
19. Wright, S. L.; Kroemer, H.; and Inada, M., J. Appl. Phys. 55, 2916 (1984); and Windhorn, T. H. and Metze, G. M., Appl. Phys. Lett. 47, 1031 (1985); and Fischer, R.; Kopp, W.; and Morkoc, H., Appl. Phys. Lett. 48, 1360 (1986).
20. Dupuis, R. D.; et al., J. Electron. Mater., in press.
21. Hensel, J. C.; Levi, A. F. J.; Tung, R. T.; and Gibson, J. M., Appl. Phys. Lett. 47, 151 (1985).
22. Smith, T. P., III; Phillips, J. M.; Augustyniak, W. M.; and Stiles, P. J., Appl. Phys. Lett. 45, 907 (1984).
23. Bevk, J.; Mannaerts, J. P.; Feldman, L. C.; Davidson, B. A.; and Ourmazd, A., Appl. Phys. Lett. 49, 286 (1986).
24. Binnig, G.; Rohrer, H.; Gerber, C.; and Weibel, E., Phys. Rev. Lett. 50, 120 (1983).
25. Golovchenko, J. A., Science 232, 48 (1986).
26. Gossman, H. J.; Bean, J. C.; Feldman, L. C.; McRae, E. G.; and Robinson, I. K., J. Vac. Sci. Technol. A3, 1633 (1985).
27. Becker, R. S.; Golovchenko, J. A.; and Swartzentruber, B. S., Phys. Rev. Lett. 54, 2678 (1985); and Phys. Rev. B32, 8455 (1985).

MATERIALS SCIENCE AND THE EVOLUTION OF INTEGRATED CIRCUIT PROCESSING

D. S. Williams, S. M. Sze, R. S. Wagner
AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, New Jersey 07974, USA

ABSTRACT

The explosion in our capability to transmit and process information is possible because of the continued improvement in cost and performance of integrated circuits. Although the minimum lithographic feature size is the processing parameter often tracked in the evolution of the technology, much of the innovation that characterizes this progress has been achieved by solving problems familiar to materials scientists.

In this discussion, the path from discrete devices to megabit memory technologies will be outlined. The evolution in fabrication processes will highlight the role of materials in the progress of this remarkable technology. The future will demand further innovation. Scaling of device geometries will continue into the 21st century as the physical dimensions to be controlled become ten's of atom layers. As a result, the challenges in integrated circuit processing will continue to be vibrant to those trained in the materials disciplines.

Introduction

The electronics industry has grown rapidly for the past two decades as a result of technological innovation in Integrated Circuit (IC) design and manufacture, and as a result of new applications for electronic systems. Electronic equipment sales in the U.S. totaled $210 billion dollars in 1986 and are projected to grow at an average rate of 12% to reach $1 trillion by the year 2000.[1] At this volume, electronic sales will exceed the sales of the automobile, chemical and steel industries. This phenomenal growth is driven at the component level by the continual improvement of IC performance and cost.

In the 1960's, the IC market was based primarily on bipolar transistors, but since 1975 the metal oxide semiconductor (MOS) transistor has grown to be the dominant device structure. The advantages of MOS in large scale IC production are derived from the processes of miniaturization. The advantages of miniaturization are twofold: increased circuit performance, and increased functional density. The circuit performance of MOS devices improves upon scaling because of the increased switching speed of the transistor itself, and because of the decrease in routing delays as transistors are more densely packed. The increase in functional density originates from the greater number of logic elements of smaller dimensions that can be placed within a given chip size. Miniaturization has resulted in a rapid succession of technologies offering advantages of dimensional scaling, reduction of power requirements and improvements in manufacturing productivity.

The performance of an MOS IC technology is often described by the product of power and propagation delay. Figure 1 shows that the power-delay product for MOS devices has decreased dramatically for the past 25 years and is projected to do so to the end of the century. Also shown in Fig. 1 is a similarly dramatic decrease in the cost per function through improvements in manufacturing productivity, even while the basic technology continued to evolve.

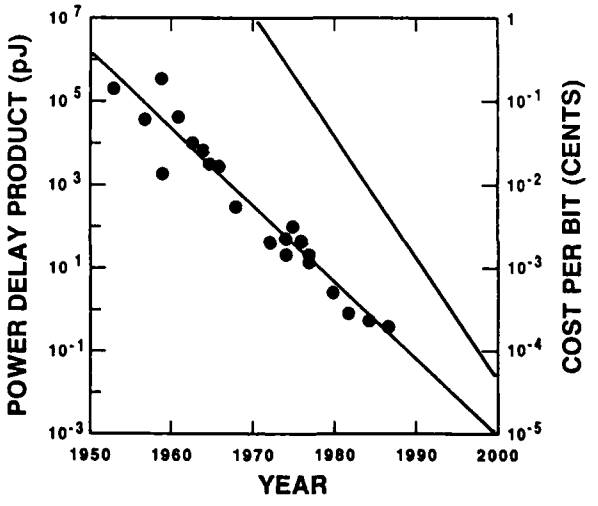


Figure 1. Power · delay product and cost per function as a function of time (i.e. the date of introduction of the IC technology.)

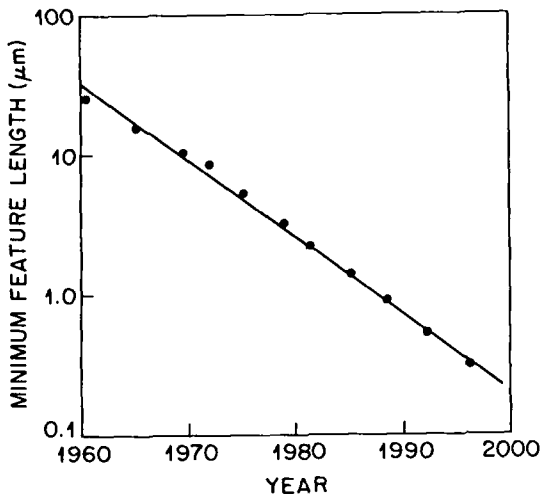


Figure 2. Minimum lithographic feature size as a function of the date of introduction of the technology.

The fabrication of integrated circuits can simplistically be described as the process of growing or depositing thin films and then selectively removing them. To define the regions of selective processing, lithographic technology is required, and the evolution of this technology has traditionally determined the pace of progress. Thus, the minimum feature size that is patterned in the integrated circuit has become a common metric to describe the evolution of MOS technology. Figure 2 shows industry progress as measured by the minimum feature size. Notice that the rate of shrinkage approximates 13% per year and has resulted in a new generation of technology approximately every 2-1/2 years. Characteristically, each new generation occupies one fourth the area and has twice the speed of the previous generation. This rate of innovation has fueled the electronics revolution, but lithographic progress does not describe the total evolutionary process. Much of the progress that has made this remarkable technology possible has required the control and characterization of processes familiar to those trained in the materials disciplines.

Device Scaling

The miniaturization of MOS devices requires the scaling of both lateral and vertical dimensions of the transistor structure as well as scaling of the operating voltages (if constant electric fields are maintained.) This reduction of transistor size has advantages in conserving "lay-out" area of the chip as well as advantages in transistor performance. Figure 3 is a schematic of an n-channel, MOS transistor (called NMOS) showing the essential features of the structure. Under normal operation, a voltage is applied to the drain while the source is grounded. No current flows between source and drain, however, unless a positive voltage is applied to the gate electrode. This gate voltage attracts electrons to the surface, thereby creating a channel beneath the gate oxide that allows current to flow from source to drain. Many such active transistors are combined to form logic elements in a digital circuit design or are combined with passive elements, such as capacitors, resistors or inductors to perform a wide range of analog circuit functions.

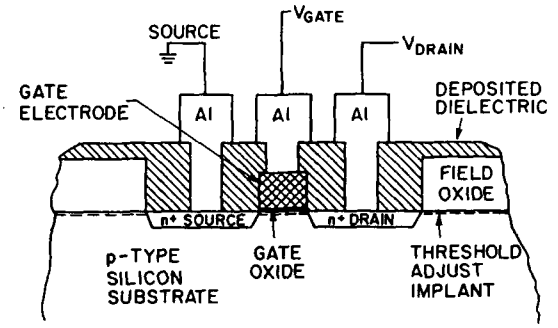


Figure 3. Schematic of an MOS device.

The impact of scaling, or the reduction of dimensions by a scaling factor $X > 1$, is summarized in Table I.[2] In scaling, the vertical dimensions of gate oxide and junction depth are decreased by a factor X^{-1}. The critical lateral dimension of gate length is also decreased by the factor X^{-1}, as is the power supply voltage for constant field scaling. The dopant concentration in the substrate increases by the factor X, but the power density on the circuit remains constant because the circuit power scales by X^{-2} while the density of devices scales as X^2. The reduction in gate length results in a decrease in the delay time of the transistor logic function itself. The result of this decrease in transit time and the decrease in power of a single device is a power-delay product that scales by the factor X^{-3}.

TABLE I

MOS Device Scaling[2]

For Constant Field Scaling by Factor of X:

Gate Length	X^{-1}
Oxide Thickness	X^{-1}
Junction Depth	X^{-1}
Supply Voltage	X^{-1}
Substrate Dopant Concentration	X^{1}
Power Density	1
Circuit Power	X^{-2}
Gate Delay	X^{-1}
Power · Delay Product	X^{-3}

Crystal Growth

The evolution of IC manufacturing practice has led to significant changes in materials and processes. Some changes were the outgrowth of scaling requirements such as the addition of refractory metal silicides to the gate electrode to decrease interconnect routing delays. However, other materials innovations have been driven solely by manufacturing productivity improvements. The art and science of crystal growth is certainly an example of productivity driven innovation. Because the cost of wafer processing is relatively insensitive to wafer size, economies of scale drive the industry to larger diameter wafers resulting in more chips per wafer.

The vast majority of silicon substrates used in IC processing are grown by the Czochralski technique. The requirements of wafers produced by these methods are that they be generally free of dislocations, while the dopant and impurity concentrations, the flatness and the thickness be rigidly controlled. Figure 4 shows the progression as increasingly larger diameter wafers have been introduced into manufacture. The figure shows that from the early days of the industry when the substrates contained only a few devices, the introduction of larger diameter of substrates has been a common event. At this writing, 150 mm diameter wafer processing lines are common, with 200 mm lines under construction. This trend toward larger wafer diameters is expected to continue through the end of the century, as developmental efforts have shown the feasibility of growth for 300 mm diameter boules.

The epitaxial growth of thin layers (<10 μm) of lightly doped silicon onto large diameter wafers is another impressively sophisticated materials process that has become commonplace in IC manufacture. The impetus for such structures is the need to dope regions around n-channel devices and p-channel devices differently. When both devices are used in logic arrays, as in Complementary MOS (CMOS) structures, epitaxial silicon with lithographically defined "tubs" is required. The material quality requirements applied to a substrate wafer are extended to epitaxial silicon growth, and crystalline defect formation in epi-wafers is a major cause of device yield degradation. Large diameter wafer requirements have spawned new designs of epi-growth reactors through the need for high throughput, lower temperature variation and robotic loading for improved cleanliness.

Silicon Oxidation

Silicon has become the dominant semiconductor substrate material for IC manufacture partly because of the properties of its oxide. High quality silicon dioxide can be thermally grown in pure oxygen or in steam as a routine manufacturing process. The thickness and integrity of this oxide is important to device performance, reliability and yield. Figure 3 shows that a thick field oxide provides isolation and a thin gate oxide controls the current flow in the transistor. As a result, the oxidation of silicon and its properties have been extensively studied and are the topic of complete symposia in support of this industry.[3] The thermal growth of gate oxide on <100> single crystal silicon is the most critical step in the entire MOS manufacturing process. Figure 5 shows the progression of gate oxide thickness as succeeding generations of technology were introduced. This gate oxide must endure electric fields of $\sim 3 \times 10^6$ volt/cm throughout the design lifetime of the device and must endure "burn-in" fields that are higher. In the development of leading edge technologies, devices with 10 nm thick gate oxides that are pin-hole free over areas of 1 cm^2 are routinely fabricated. Furthermore, techniques for growing <10 nm oxides are under development. However, many important processing details remain in growing oxides of the thicknesses required that can sustain such high electric fields. Figure 6 is a high resolution TEM micrograph of thick SiO_2 thermally grown on <100> silicon after improper cleaning procedures. Notice the non-uniformity of the SiO_2/Si interface, which will lead to non-uniform electric field distributions and thus to premature breakdown of the oxide. The non-uniformities of the interface in Fig. 6 are a significant fraction of the total oxide thickness that will be required by device scaling in mid-1990. Clearly many challenges in the field of oxidation of silicon will be encountered as scaling of device dimensions continues.

Gate Materials

The gate acts as the logic switch in the MOS device by generating a conducting region beneath the gate oxide that allows current to flow from source to drain. The gate electrode length is usually fabricated as the minimum lithographic dimensional, and requires strict dimension control during processing because the device switching speed is directly related to length. Figure 7 shows the selection of gate materials through the evolution of MOS technology. Aluminum gate electrodes were used in the early days of IC fabrication when p-channel transistors dominated the technology (called PMOS). However, the invention of the "self-aligned" polysilicon gate electrode coincided with the conversion to n-type MOS transistors (NMOS) and has been one of the most enduring processing innovations.

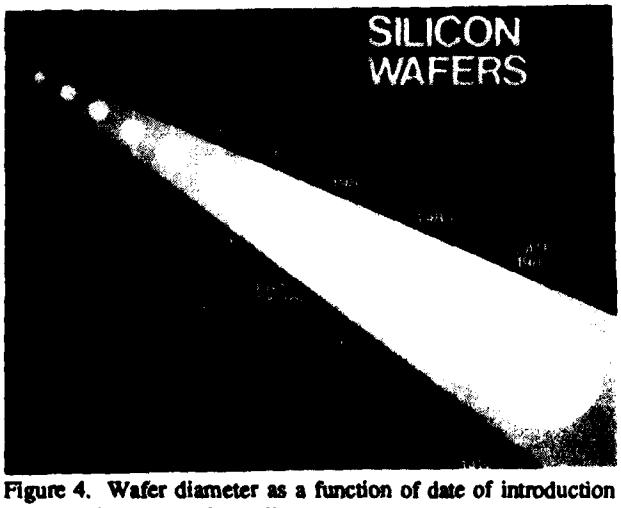


Figure 4. Wafer diameter as a function of date of introduction of the technology. (From K. E. Benson et al., unpublished).

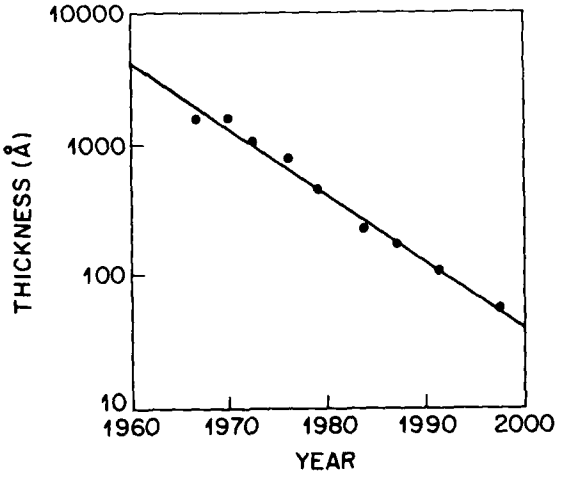


Figure 5. Thickness of gate oxide as a function of the date of introduction of the technology.

In the early years of MOS device processing, the junctions in single crystal silicon which form the source/drain areas were formed by diffusion of dopants into the substrate at high temperatures (1000-1100°C) followed by the alignment of the aluminum gate electrode to the diffused regions. A major advance in manufacturing throughput and yield of MOS structures occurred with the selection of a gate material of "polysilicon." The chemical vapor deposition of polycrystalline silicon (polysilicon) onto the gate oxide, which was followed by a gate electrode patterning step, allowed the gate to be used as the "self-aligned" mask for ion implantation or diffusion of source/drain dopants into the substrate. This change of material enabled the elimination of the gate misalignment tolerances from the transistor specifications, thereby greatly improving performance and device yield.

This change in gate electrode material did not come without a penalty, however. Aluminum has a resistivity of 3×10^{-6} Ω–cm whereas heavily doped polysilicon has a resistivity of $\sim 5 \times 10^{-3}$ Ω–cm. When device designs use the gate electrode to route signals long distances, such as in large memory circuits, the propagation delay of the gate electrode can adversely affect circuit performance. Figure 8 shows the impact of gate routing delays on circuit performance for differing interconnect materials.[4] For polysilicon gates, the delay in routing a signal 1 cm is several orders of magnitude greater, even at 10 μm design rules, than the MOSFET delay.

An improvement in gate interconnect delay that has been widely adopted by the MOS industry involves yet another material change, or rather a material addition. By "strapping" a refractory metal silicide onto the self-aligned polysilicon gate material, the advantages and experience base of the self-aligned polysilicon gate can be retained while the gate interconnect resistivity can be decreased by a factor of ~10 (Fig. 8). The silicon saturated refractory metal silicides of Mo, W, Ta and Ti were carefully scrutinized for processing related properties such as high temperature stability, deposition process and stoichiometry control, oxidation rate, wet and dry etch rates and resistivity.[5] Figure 9 shows a TEM cross section of a $TaSi_2$/polysilicon gate electrode with junction delineation showing the n^+/p source/drain profiles. This gate electrode structure was introduced into manufacture by AT&T-Microelectronics in a 256K DRAM process in 1981 and remains in large volume manufacture today.

Scaling of device geometry continues to pressure interconnect delay considerations. Figure 8 shows that even with aluminum metallization, interconnect delay will limit logic function performance at sub-micron design rules. In addition, junction depths have become shallower and with smaller window sizes, thus driving the contact resistance of metal to source/drain region higher. Recent processing developments that decrease these parasitic resistivities use a process to form a "self-aligned" silicide in the source, drain and gate. The method of such metallization is illustrated in Fig. 10 where regions of silicon dioxide are patterned to expose substrate silicon or patterned polysilicon. Metal is then deposited and the structure annealed to react metal with the silicon substrate and with the polysilicon gate to form a metal silicide. A selective wet chemical etch then removes the unreacted metal leaving the metal silicide in selective locations. Such processes

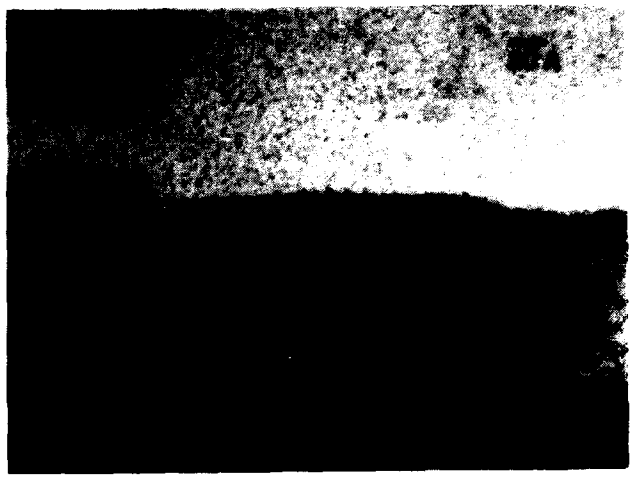


Figure 6. High resolution TEM cross-section of the interface between a thermal oxide grown on <100> silicon single crystal. (From J. M. Brown, unpublished).

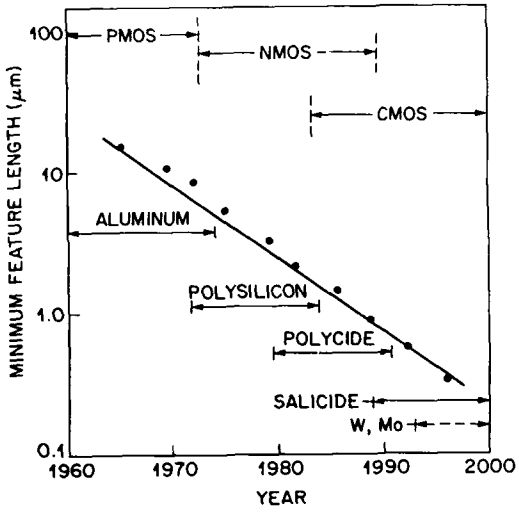


Figure 7. Gate material and transistor type as a function of date of introduction of the technology.

are well developed for the formation of titanium disilicide and cobalt disilicide. Sub-micron design rule technologies will use such interconnect schemes to enhance the performance of MOS devices through reduced contact resistances and smaller routing delays.

The future of gate electrode technologies may yet lead to other material changes. The refractory metals tungsten and molybdenum are under study as replacements for the polysilicon gate material. Both materials have lower resistivities than polysilicon, have excellent etch properties for patterning sub-micron features, and have appropriate work functions for application to low voltage CMOS processing. Gate electrode processing continues to be an exciting and controversial area of IC process development.

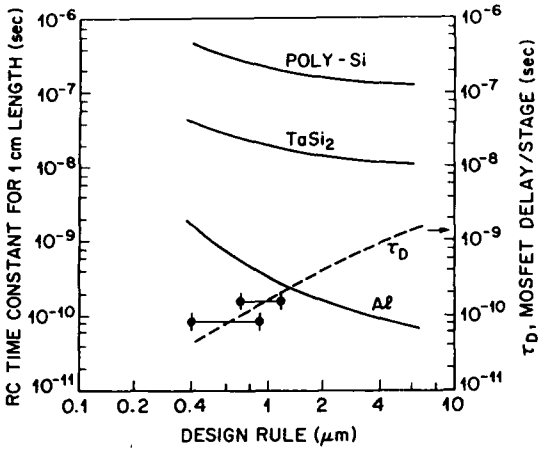


Figure 8. The impact of gate electrode material on interconnect routing delays as a function of design rules. Doped polysilicon (30 Ω/□), $TaSi_2$/Polysilicon (2.5 Ω/□) and Al (0.025 Ω/□) with 1 μm thick dielectric overlayers on 1 μm thick field oxide. (From Sinha et al.[4]).

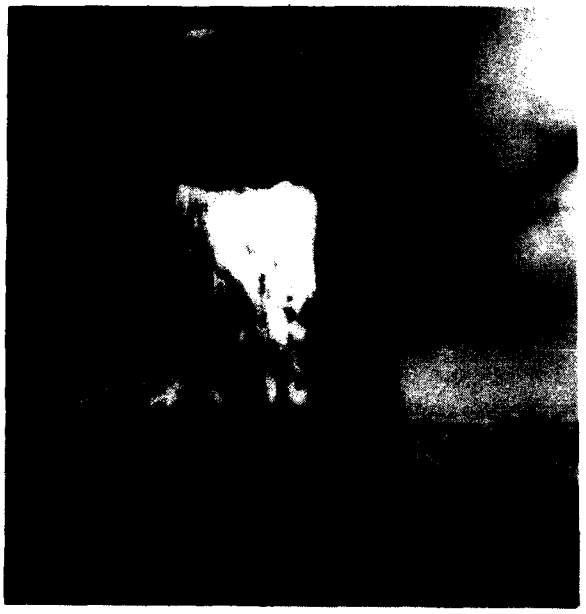


Figure 9. A TEM cross-section of a $TaSi_2$/Polysilicon gate structure. Junction staining techniques highlight the heavily doped region of the source/drain. (From T. T. Sheng, unpublished).

Dopants in Silicon and Silicides

The control of the depth of dopant diffusion in the formation of junctions has been an important aspect in IC processing from its inception. Whether forming emitter/base/collector structures in bipolar transistors or source/drain junctions in MOS transistors, control and reproducibility of dopant diffusion has been an important portion of the technology. The diffusion of dopants from gaseous sources (e.g. PBr_3, $POCl_3$, BBr_3) in hot walled tube furnaces at high temperatures was satisfactory in the early years

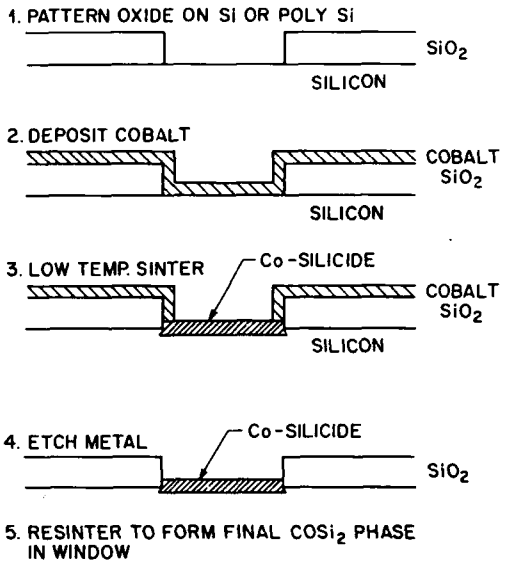


Figure 10. Processing sequence for the formation of a self-aligned silicide.

of IC manufacture. As total dose control became more critical and diffusion depth became more shallow, ion implantation began to replace furnace diffusion in critical stages of device processing. Currently, to set substrate "tub" doping to depths of several microns in the fabrication of CMOS devices, furnace diffusion of dopants is still common. However, the formation of junctions in that tub structure is done by ion implantation. The annealing time-temperature sequence is selected to repair radiation damage and to electrically "activate" the dopant but not to alter the diffused substrate dopant profile.

Because scaling to submicron design rules requires the scaling of junction depths as well as the incorporation of silicides into the substrate region, the behavior of dopants in the self-aligned silicide structures is an active area of materials research. The solubility, diffusivity and segregation behavior of dopants in the silicide/silicon systems is poorly understood and critically important to continued device scaling. As an example of the complex interplay between dopants, metal-silicide and silicon, consider the formation of shallow junctions (<100 nm) beneath a self-aligned silicide. As stated, scaling of device geometry requires scaling of junction depths. At submicron design rules, the junction depths required become less than 0.25 μm, and if siliciding of the source/drain area is employed, the silicide thickness plus the junction depth must fall within this maximum depth. The formation sequence of such shallow junctions is further complicated by the segregation behavior of dopants during the silicide formation process. Schemes have been proposed that implant dopant into silicon before silicidation, into or through metal prior to silicidation, into or through intermediate silicides, and into or through the final disilicide. Obviously, these options include a large number of experimental parameters and the problem is made more complex by the lack of fundamental diffusion, solubility and segregation data. One successful method that produces shallow junctions beneath $CoSi_2$ is to implant dopant into the disilicide.[6] The implant energy is selected to contain

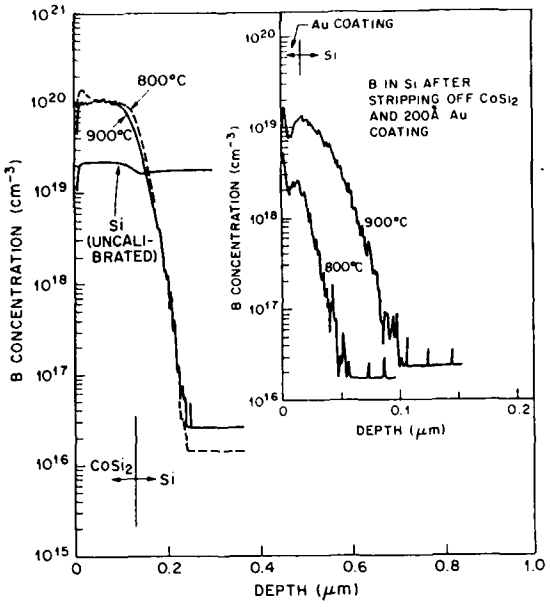


Figure 11. SIMS profile of boron beneath 1000 Å $CoSi_2$. After chemical strip of the $CoSi_2$ (insert) the interface concentration and junction depth can be determined as a function of anneal temperature. (From Liu et al.[6]).

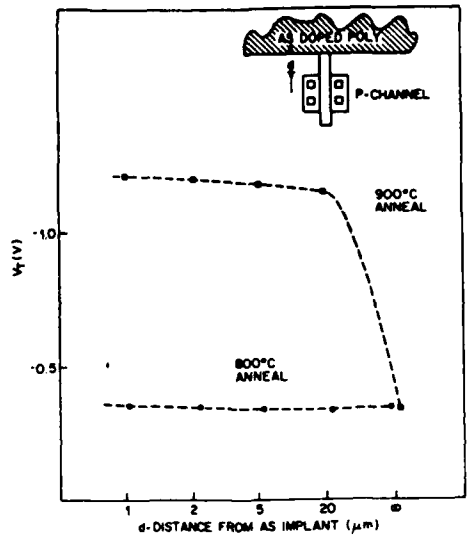


Figure 12. Threshold voltage shifts of a p-channel device resulting from lateral diffusion of arsenic through the $TaSi_2$ and from segregation of boron to the $TaSi_2$. (From S. J. Hillenius et al.[7]).

the implantation damage within the silicide and the dose is selected to exceed the solubility limit of dopants in the silicide. The subsequent anneal can be chosen at a temperature that diffuses dopant into the underlying silicon, but not at the higher temperature necessary to anneal radiation damage. Figure 11 is a Secondary Ion Mass Spectrometry (SIMS) profile of boron difluoride implanted into $CoSi_2$ in the described fashion. The inserted SIMS profiles following wet chemical removal of the $CoSi_2$ show accurate junction depths and estimates of interface concentrations. Note that the junction depth is 50-100 nm beneath the silicide/silicon interface. Such junction formation techniques will satisfy scaling requirements well into the next decade.

As a second example of new challenges in dopant-silicide interactions, Fig. 12 shows the threshold voltage behavior of a p-channel device that is electrically connected to an n-doped region by a $TaSi_2$/Si gate electrode.[7] This is a structure common to CMOS IC's and represents a "worst-case" geometry with respect to relative area. The dependent variable plotted in Fig. 12 is the threshold voltage of the p-channel transistor positioned from the n-doped region by a distance measured in microns. Note that the control in this experiment is a p-channel transistor completely separated with respect to the n-doped region (distance ∞). Figure 12 shows that for 900°C-30 min anneals, the threshold voltage of the p-channel MOSFET shows the influence of the diffusion of n-type dopant up to a distance of 20 microns. This effect on the threshold voltage, which is shifted to more n-type polysilicon behavior, is the combined effect of the segregation of boron dopant from the p-channel polysilicon gate to the silicide, as well as the high rate of lateral diffusion of n-type dopant in the silicide. Estimates of the dopant diffusivity in silicides suggest several orders of magnitude higher values than in polysilicon but no quantitative studies have been undertaken.

As these examples show, the use of dopants in future MOS technology will not be a dormant field, as scaling of geometry and the introduction of new materials create unexplored processing regimes. Furthermore, the demand for reliable data on the fundamental material properties of dopants in metal/silicide/silicon systems is apparent.

Interconnect Metallization

The interconnection of logic functions by high conductivity materials has been a stable and robust technology since the early manufacture of MOS IC's. Although gold interconnection schemes are used by some, aluminum interconnect metallurgy has dominated. The advantages of aluminum over gold are its relative ease of patterning, its compatibility with silicon substrate electrical properties, and its adhesion to SiO_2. The early history of MOS manufacture used only one level of aluminum interconnection. The increased functional density and the advent of more sophisticated CAD tools has led to multiple levels of metal. Now three levels of aluminum are becoming commonplace and demands from the design community for more levels are driven by the custom and semi-custom IC market.

Aluminum runners of near-minimum lithographic design rules have been common throughout this rapidly changing technology. Most aluminum alloys contain 0.5-1.0% silicon to satisfy the solubility of the metal and to avoid substrate dissolution at the annealing temperatures (called junction "spiking"). The alloys may also contain up to 4.0% copper for improved electromigration resistance. The current densities required of these runners are up to 5×10^5 Amp/cm^2 and few electromigration failures are observed when existing design guidelines are followed. However, submicron design rules are creating conditions that will require the re-examination of the standard aluminum metallization practices.

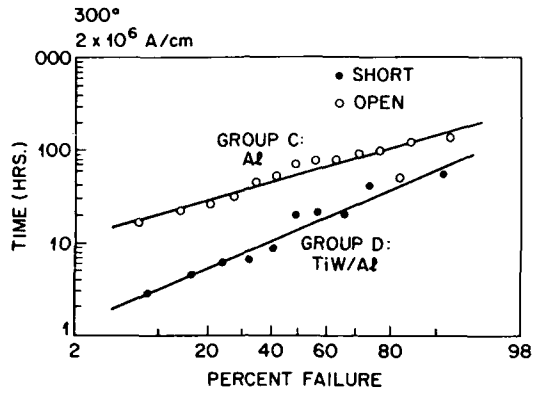


Figure 13. Short circuit failure mode that becomes dominant in multiple level metallization schemes. Changes in failure distribution from open circuit to short circuit failure mode upon introduction of Ti:W barrier layer. (From J. M. Towner[8]).

Scaling of MOS device geometries has drastically reduced lateral dimensions in most layout schemes, but has allowed some vertical dimensions to scale more slowly. The results of this practice are severe topographies that lead to non-uniform deposition of aluminum because of self-shadowing during the physical vapor deposition process. The questionable reliability of aluminum as it relates to the ability to cover topography at submicron design rules is driving the need for innovation in metallization technology. Improvements in aluminum deposition technology as well as chemical vapor deposition of tungsten are competing to satisfy this pressing need for better step-coverage. The evolution of metallization deposition technology during the next decade is assured.

The use of multiple levels of metal has also led to the re-examination of failure modes in electromigration experiments. Towner and co-workers[8] have shown that the mechanisms active in determining the mean time to failure (MTF) under accelerated testing at high current densities can be changed from the traditional open circuit electromigration failure to short circuit failure. Open circuit failure is the result of material transport by electromigration at high current densities until a local instability is reached and a complete break in the line occurs. Short circuit failure is the growth of aluminum "whiskers" or "hillocks" during high current stressing until one runner "shorts" to a neighboring runner. Figure 13 shows the impact of adding a thin Ti:W alloy layer beneath the aluminum runner. This "barrier" film is used to prevent metallurgical interaction between aluminum and the silicon substrate to avoid junction "spiking". The outgrowths from the original patterned aluminum features, which are enhanced by the presence of a refractory barrier layer, can short circuit to neighboring lines. In multiple level metal schemes, short circuits occur either in the same level or to routing levels above or below. Many of the proposed solutions to the problems of topography coverage and of contact stability exacerbate the short circuit failure mode.

The next decade will require adjustments to traditional aluminum metallization practice driven by coverage of the severe topography generated at submicron design rules. However, alterations to conventional practice can require a complete re-evaluation of reliability criteria because of changes in the materials properties that limit device lifetime. Clearly metallization of MOS structures will be a dynamic field through the next decade.

Technological Trends

The future will demand continued innovation in materials and processing for IC device technology. As Figs. 1 and 2 show, gains in circuit performance and logic density will be realized at design rules well below 0.5 micron. Therefore scaling of traditional device geometries should continue into the next century even for conventional bipolar and MOS device technologies. Although silicon should continue to be the dominant material, the future suggests the incorporation of heterostructures, multilayer structures and combinations of devices on the same chip. This strategy is driven by the desire to make integrated photonic and electronic designs on the same device, using photonic interconnects and combinations of bipolar, GaAs or MOS logic. All scenarios point the way to higher functional density, greater complexity of materials and more stringent requirements of material chemistry and microstructure. Of course, productivity will continue to benefit from larger diameter substrates, and the finer design rules will favor the process control and the uniformity afforded by single wafer processing. As the complexity of devices and structures increases, control of structural dimensions, dopant and impurity concentrations and microstructures will guarantee fascinating challenges to the best trained materials scientists and engineers.

REFERENCES

1. S. M. Sze, Ed., *VLSI Technology*, 2nd Edition, McGraw-Hill, New York, (1988).
2. J. D. Meindl, Solid State Techn. 30(12), 85 (1987).
3. J. J. Simonne and J. Buxo, Eds., Proceedings 4th Conf. on Insulating Films on Semiconductors, North Holland, New York, (1986).
4. A. K. Sinha, J. A. Cooper and H. J. Levinstein, IEEE Electron Dev. Lett. *3(4)*, 90 (1982).
5. S. P. Murarka, *Silicides for VLSI Applications*, Academic Press, New York, (1983).
6. R. Liu, D. S. Williams and W. T. Lynch, J. Appl. Phys. *63(6)*, 1990 (1988).
7. S. J. Hillenius, R. Liu, G. E. Georgiouu, R. L. Field, D. S. Williams, A. Kornblit, D. M. Boulin, R. L. Johnston and W. T. Lynch, IEDM Tech. Digest, 252 (1986).
8. J. M. Towner, Proc. 23rd Reliability Physics Symposium, 81 (1985).

BULK CRYSTAL GROWTH OF III-V AND OTHER MATERIALS FOR OPTICAL COMMUNICATION BEYOND 1.5 μM

R. A. Laudise, W. A. Sunder, R. L. Barns
AT&T Bell Laboratories
Murray Hill, New Jersey 07974, USA

J. M. Parsey, Jr.
AT&T Bell Laboratories
Reading, Pennsylvania 19604, USA

P. M. Bridenbaugh
AT&T Bell Laboratories
Holmdel, New Jersey 07733, USA

BULK CRYSTAL GROWTH OF III-V AND OTHER MATERIALS FOR OPTICAL COMMUNICATION BEYOND 1.5μm

R. A. Laudise, W. A. Sunder, R. L. Barns

AT&T Bell Laboratories
600 Mountain Avenue
Murray Hill, New Jersey 07974

J. M. Parsey, Jr.

AT&T Bell Laboratories
2525 North 12th Street
Reading, Pennsylvania 19604

P. M. Bridenbaugh

AT&T Bell Laboratories
Crawfords Corner Road
Holmdel, New Jersey 07733

ABSTRACT

Optical communication has provided a driving force for much of electronic materials research and development. As optical fibers improve and losses approach the Rayleigh scattering limit $\propto 1/\lambda^4$, fibers and their associated devices will move to longer wavelengths. Heterostructure lasers will require lattice matched single crystal substrates for the regime 1.5μ to 5μm. This paper reviews the materials choices, growth and present status of such substrates. Materials parameters are examined and shown to suggest single crystal GaSb for $\lambda > 2$ μm and (CdMn)Te for longer wavelengths. Conditions for Czochralski growth of GaSb are discussed and preliminary results on the partition of dopants are described. Conditions for the Bridgman growth of (CdMn)Te are described and the distribution constant of Mn is shown to be close enough to one that most of a grown boule has its lattice parameter matched satisfactorily for good epitaxial growth of (HgCd)Te. Mn is shown to lattice harden (CdMn)Te and repress twinning.

OPTICAL COMMUNICATIONS needs have provided a driving force for much research on electronic materials ever since the earliest days of the laser when it was realized that communication on an optical frequency carrier provided the possibilities of orders of magnitude improvement in band width. As optical fibers improved (Figure 1)[1][2], it became apparent that as the loss approached the Rayleigh scattering limit, where loss $\propto 1/\lambda^4$, longer wavelength lasers and detectors would be needed.[1]

At first wavelengths were chosen to avoid absorption due to OH, but as glass purity improved the maximum useable wavelength tended to be set only by the onset of lattice absorption in SiO_2. Hence there was a progression from 0.83 to 1.3 to 1.5μm in systems based on silica glass. Presently research is active in many laboratories on glasses whose lattice absorption is at wavelengths beyond that of SiO_2, so that even lower Rayleigh scattering will occur. Thus there is an interest in materials whose band gaps will provide lasers and detectors at and beyond 1.5μm. In order to make devices, multi-layer lattice matched structures must be deposited on single crystal substrates. As can be seen from Figure 2, [1][3] the deposition of III-V solutions in the (InGa) (AsSb) system on GaSb substrates will make accessible the wavelength range up to approximately 4.5μm. In addition, due to spin orbit valence band splitting, GaSb based detectors have have been demonstrated to have increased sensitivity (superior to InP based devices) at 1.3μm.[4] Consequently, the growth[5] of high quality single crystal GaSb b comes of importance. Small band gap II-VIs will make accessible the wavelength range beyond 1.5μm out to wavelengths well beyond 5μm. Typical band gap and lattice parameters are given in Figure 3.[6] Epitaxially deposited (HgCd)Te approximately lattice matched to CdTe single crystal substrates is already used as an infrared sensor. It is the material of choice for infrared detector arrays for the 3-5μm and the 10-14μm atmospheric windows and shows promise in the 1-2μm optical communication range. Bridenbaugh has recently shown that single crystal (CdMn)Te can lattice match (HgCd)Te over the wavelength range 1 to 20μm, so that it should be an attractive substrate material for detectors and lasers in that region.[7] Spin orbit valence band splitting should also improve detector sensitivity in this system. In this brief paper we summarize some of our recent results on the preparation of GaSb[5] and (CdMn)Te.[7]

GaSb - The melting point of GaSb is 705-710°C with the Sb vapor pressure approximately 10^{-2} Torr.[8][9] Although Czochralski growth has been reported many times in the literature,[10]-[14] impurity distribution, interface shape, perfection and growth conditions have not been correlated in detail. Recently we completed such a study,[5] part of which is reported here. Our apparatus[15] and growth parameters are given in Table 1.

GaSb was synthesized by melting together Ga and Sb in a SiO_2 crucible and pulling a polycrystalline boule. A slight oxide film was sand blasted and etched from the boule and it was re-melted to grow the crystal. Typical of the doping results obtained are those for Te shown in Figure 4. Carrier concentration versus Te analysis studies[5] show that background acceptor concentration is probably approximately 5×10^{17}. Spectrochemical analysis suggests that acceptors such as Si and Mg are present in this range. In addition, anti-site defects where As occupies a Ga site (As_{Ga}) leading to n-type carrier concentration of approximately 10^{16} have been reported in GaAs[16] and might be expected in GaSb. In 1967 Van Der Meulen[17] postulated an acceptor center $Ga_{Sb}\ V_{Ga}$ to explain the high accepter concentration ($\approx 10^{17}$) found in purified undoped GaSb. In our experiments we find undoped GaSb to be p-type in the range $1 - 3 \times 10^{17}$ so that we expect this defect is present in our material. Above approximately 10^{19} carrier concentration saturates probably because of Te complexing. Similar results were obtained with Se. Ge doping results in p-type conductivity in the range $10^{17} - 10^{19}$. Approximate distribution constant results are given in Table 2. However, the simple normal freeze (Scheil) equation,

$$\log c/c_o = \log k + (k-1) \log (1-g) \qquad (1)$$

where c is the impurity (dopant) concentration in the solid, c_o is the impurity concentration in the liquid, k is the distribution contact and g is the fraction of the melt frozen is not obeyed. However, if the normal freeze equation is corrected for changes in Ga/Sb in the melt as freezing goes on, with the assumption that GaSb does not melt congruently a better fit is obtained. Hurle, et al.[18][19] first pointed out this effect in InP. In the present case, the law of mass action dictates for the reaction:

$$GaSb_s \rightleftarrows Ga_l + V_{Sb} + 1/n\ Sb_{n\ g} \qquad (2)$$

that

$$K_2 = V_{Sb} p_{Sb_g}^{1/n} \qquad (3)$$

where V_{Sb} is a vacancy at the Sb site. The incorporation of Te follows the reaction

$$Te_l + V_{Sb} \rightleftarrows Te_s \qquad (4)$$

so that

$$k_{Te} = \frac{(Te_s)}{(Te)_l V_{Sb}} \qquad (5)$$

Thus k_{Te} is a function of Ga/Sb.

The modified Scheil equation which results is derived by Sunder et al.[5] Further complications arise because interface shape is not flat so that facets form and interface shape changes as growth proceeds. However, in spite of these difficulties low dislocation (approximately $10^2-10^3\ cm^{-2}$) material has been grown at high doping levels ($> 10^{18}\ cm^{-3}$). The presumed mechanism is lattice hardening as has been previously reported for other III-Vs.[20] The most important remaining challenge for GaSb is the preparation of high resistivity material.

(CdMn)Te - Bridenbaugh[7] has shown that (CdMn)Te can be lattice matched to (HgCd)Te over the range where band gap corresponds to wavelengths from 1 to beyond 20μm.[6] Bill and Sen[21] have reported the use of $(Cd_{.96}Zn_{.04})Te$ substrate to lattice match $Hg_{0.8}Cd_{0.2}Te$. However, the distribution coefficient for Zn is ≈ 1.5 so that large fractions of grown boule are not lattice matched to the desired composition of the HgCdTe epi-layers. (CdMn)Te has been grown by the Bridgman technique under the conditions shown in Table 3, and the distribution coefficient for Mn is closer to unity.

The tendency to twin is negligible in comparison to the high incidence of twinning in CdTe. The greater the Mn content, the lower the incidence of twinning presumably because stacking fault energy is reduced.[22] When Mn exceeded 10 mol%, twinning was rare, and even below 10 mol% twinning was reduced significantly from that found in CdTe produced under similar conditions. Indentation tests show that Mn hardened the lattice. Since deformation can occur by twinning through stacking faults essentially by the shearing motion of one layer over another, it is reasonable that Mn substitution in a layer can inhibit the shearing movement. In addition the larger band gap increases the restivity of Mn doped material. The distribution coefficient for Mn in $Cd_{.9}Mn_{.1}Te$ has been determined as 0.966 ± 0.005[7]. The distribution constant does not vary measurably with Mn so that the majority of a boule will lattice match to a particular epitaxial layer within 0.04%, a rule of thumb often used to predict good epitaxy. Using the normal freeze equation, the lattice parameters of the end members, Vegard's law and the experimental distribution coefficient, Table 4 was generated. Mn contents up to 0.1 will lattice match (HgCd)Te over the composition range of interest for communications and infrared applications.

CONCLUSION

The first step in opening a new laser and detector wavelength regime is developing reproducible substrate material for the needed heteroepitaxial structures. It appears that for the wavelength range 1.3-4.5μ this step has been achieved with GaSb single crystals of the required size and perfection now available. (CdMn)Te fulfills similar requirements for the range from 1.0μ to well beyond 5μm. Already GaSb laser detectors[4] and (GaIn) (AsSb)-(GaAl) (AsSb) based double heterostructure lasers which have been optically pumped and lase at $\approx$ 2μm have been made. In addition, this system has been shown to lase at room temperature at 1.6μm.[23] Herman et al[24] describe (HgCd)Te/CdTe heterostructure properties and applications. A most intriguing recent use of CdTe is in an ultra-fast (picosec) detector prepared by UV infrared MOCVD epitaxy of polycrystalline CdTe on glass[25]

REFERENCES

(1) R. A. Laudise, *J. Cryst. Gr. 65,* (1983), 3-23, and references contained therein.

(2) W. G. French, J. B. MacChesney, P. B. O'Connor and G. W. Tasker, Bell Sys. Tech. J. *53* (1974) 951.

(3) Courtesy of P. K. Tien.

(4) O. Hilderbrand, W. Kuebart, K. W. Bentz, M. H. Pilkuhn, *IEEE J. Quantum Electron,* QE-17, (1981), 284.

(5) W. A. Sunder, R. L. Barns, T. Y. Kometani, J. M. Parsey,Jr. and R. A. Laudise, *J. Cryst. Gr., 78* (1986) 9.

(6) "Epitaxial Growth of II-VI Semiconductors for Telecommunications," R. D. Feldman, D W. Kisker, P. M. Bridenbaugh, R. F. Austin, and K. F. Jeffers, Advanced Materials for Telecommunications, Editiors - P. A. Glasow, Y. I. Nissin, J. P. Noblanc, J. Speight 191, (1986) by LES Editions of Physique.

(7) Bridenbaugh, Materials Letters, Vol. 3 p. 287 (1985).

(8) M. Hansen, Ed., *Constitution of Binary Alloys,* (McGraw Hill, New York, 1958) p. 755.

(9) D. Effer and P. J. Etter, *J. Phys. Chem. Solids 25* (1964) 451.

(10) W. P. Allred in *Compound Semiconductors,* Vol. I, Preparation of III-V Compounds, eds. R. K. Willardson and H. L. Goering, p. 187-193, New York: Reinhold Publishing Corporation, 1962.

(11) M. Kumagawa, H. Oka, N. V. Quang and Y. Hayakawa, *Jpn. J. Appl. Phys. 19* 4 (1980) 753-754.

(12) S. Tohno and A. Katsui, *J. Electrochem. Soc. 128* (1981) 1614.

(13) B. Cockayne, V. W. Steward, G. T. Brown, W. R. MacEwan and M. L. Young, *J. Cryst. Gr. 58* (1982) 267-272.

(14) S. Kondo and S. Miyazawa, *J. Cryst. Gr. 56* (1982) 39-44.

(15) E. Buehler, *J. Cryst. Gr. 43* (1978) 584.

(16) D. M. Hofmann, B. K. Meyer, F. Lohse and J. M. Spaeth, *Phys, Rev. Lett. 53* (12), (1984), 1187.

(17) Y. J. Van Der Meulen, *J. Phys. Chem. Solids 28,* (1967), 25.

(18) D. T. J. Hurle, private communication.

(19) M. R. Brozel, E. J. Foulkes, I. R. Grant, L. Li, D. T. J. Hurle, and R. M. Ware, *J. Cryst. Gr. 70* (1984) 191.

(20) Y. Seki, J. Matsui and H. Watanabe, *J. Appl. Phys. 47 (1976) 3374; 49 (1978) 822; see also B. Cockayne, G. T. Brown and W. R. MacEwan, J. Crys. Gr. 54* (1981) 9, and references contained therein.

(21) S. L. Bill and S. Sen, *J. Vac. Sci. Tech. A31,* (1985), 112.

(22) H. Gottschanek, G. Patzer and H. Alexander, *Phys. Stat. Sol. (a)45,* (1985), 207.

(23) N. Kobayashi, Y. Horikoshi, C. Uemura, Japn. J. Appl. Phys., 1980, *19,* pp. L30-L32.

(24) M. A. Herman, M. Pesa, J. Appl. Phys., Vol 57, *8,* 1985, p. 2671.

(25) A. M. Johnson, D. W. Kisker, W. M. Simpson, and R. D. Feldman, Proceedings of the Picosecond Electronics and Optoelectronics Conference, Lake Tahoe, Nev, Mar. 13-15, 1985, Pub. by Springer Verlag. Ed. by G. A. Mourou, and D. M. Bloom and C. H. Lee, p. 188.

FIGURE CAPTIONS

Figure 1. Improvements in Loss-Wavelength Characteristics of Multimode SiO_2-GeO_2 Optical Communications Fibers.[1][2]

Figure 2. Lattice Parameter and Band Gap of III-Vs.[1][3]

Figure 3. Lattice Parameter and Band Gap of II-VIs.[6]

Figure 4. Carrier Concentration versus Mobility of GaSb.

TABLE 1

GROWTH CONDITIONS - GaSb

Apparatus	- Czochralski Crystal Puller devised by Buehler[15] using 20 kw 450 kHz r.f. generator
Crucible	- Vitreous SiO_2 (and in a few cases BN)
Sb/Ga (mole ratio)	- (made Sb rich to compensate for Sb volatilization) = 1.001
Rotation Speed	- 50-55 rpm
Pull Rate	- 0.48 in/hr. (1.2 cm/hr.)
Crystal Orientation	- <111> B (Sb)
Atmosphere	- H_2, p = 10.5 psi gauge, flow rate = 45 cc/min.

TABLE 2

APPROXIMATE DISTRIBUTION COEFFICIENTS FOR GaSb

Te ≈	0.4	(donor)
Se ≈	0.2	(donor)
Ge ≈	0.1	(acceptor)

TABLE 3

GROWTH CONDITIONS FOR (CdMn)Te

Apparatus	-	Vertical Resistively Heated Bridgman Furnace
Crucible	-	Vitreous SiO_2 capsule coated on the inside with pyrolytic graphite evacuated to 10^{-4} Torr.
Growth Rate (by lowering crucible)	-	2 mm/hr.
Temperature Gradient	-	10-20°C/cm. at the growth interface

TABLE 4

USEABLE FRACTION of CdMnTe BOULE

CRYSTAL COMPOSITION	LATTICE PARAMETER	FRACTION OF BOULE WITH LATTICE PARAMETER CONSTANT TO .04%
$Cd_{.9}Mn_{.1}Te$	6.4649 Å	.99
$Cd_{.7}Mn_{.3}Te$	6.4327 Å	.80
$Cd_{.5}Mn_{.5}Te$	6.4005 Å	.62

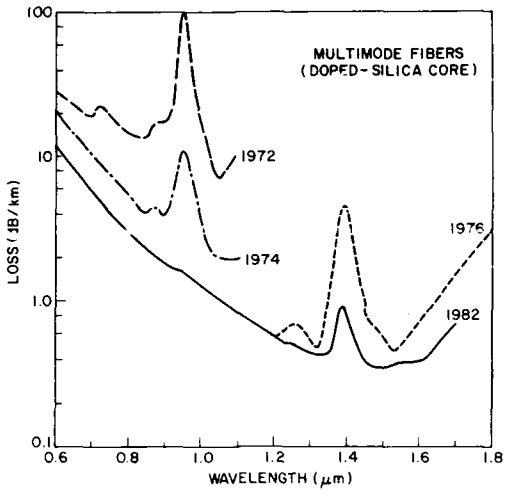


Figure 1

Improvements in Loss-Wavelength Characteristics of Multimode SiO_2-GeO_2 Optical Communications Fibers.[1] [2]

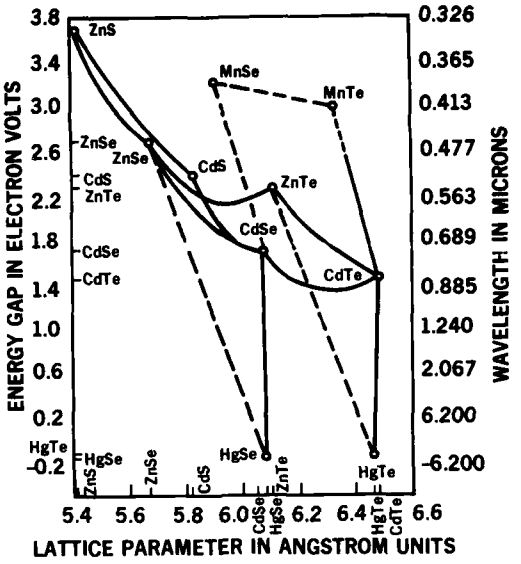


Figure 3

Lattice Parameter and Band Gap of II-VIs.[6]

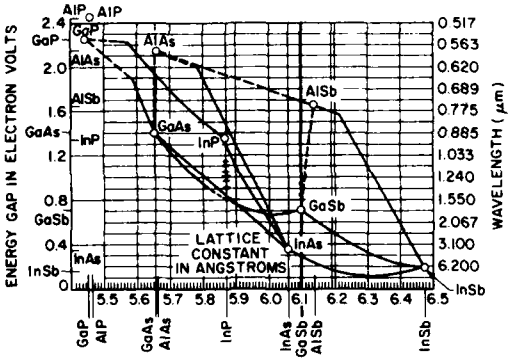


Figure 2

Lattice Parameter and Band Gap of III-Vs.[1] [3]

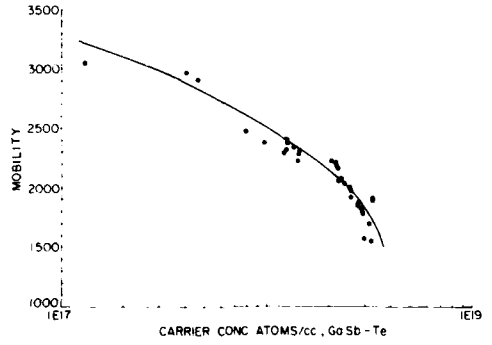


Figure 4

Carrier Concentration versus Mobility of GaSb.

MATERIALS ASPECTS OF SEMICONDUCTOR PROCESSING

C. Sheldon Roberts
Consultant, Materials and Processes
Sunriver, Oregon 97707, USA

ABSTRACT

A very large percentage of semiconductor integrated circuits (IC's) are made from silicon single crystal wafers of the Czochralski-grown (CZ) variety by complex processes such as epitaxial growth, diffusion, ion implantation, rapid thermal annealing and metallization. The rather simple diffused transistor system of the early '60's (Si, SiO_2 by oxidation, Al and Au-Si eutectic) has evolved to the "grand composite" IC of the '80's (Si, epi-Si, SiO_2 by several methods, Si_3N_4, polysilicon, silicides, refractory metals, Al alloys, PSG and organic overcoats) featuring in situ materials engineering and multilayer interconnects. Materials characteristics and process principles are virtually inseparable in this technology. This is also true of the silicon-on-insulator (SOI) processes, which are discussed.

THE INTEGRATED CIRCUITS of greatest volume in current manufacture fall in the VLSI (Very Large Scale Integration) category, containing 65,000 - 2,000,000 components per chip. For example, a 1 Megabyte DRAM (Dynamic Random Access Memory) of which many versions are now in production, range from 7-8 mm square and have feature sizes between 1 and 1.5 μm. The dominant technology for digital VLSI has been NMOS, where arrays of n-channel MOSFET's (Metal Oxide Semiconductor Field Effect Transistors) similar to the simple element shown in Fig. 1, turn each other on and off at such great speeds that access times are of the order of 10's of nanoseconds.

A dozen different materials are used in these circuits, with lateral dimensions in the μm range and vertical dimensions less than 1 μm. In the case of the gate oxide of a MOSFET, vertical thickness dimensions are now in the region 10-30 nm (100-300A). Using 10-15 fabrication steps, manufacturers are making a composite structure of greater complexity and finer dimension than has yet to be attained in the field of load-bearing composites.

The n-channel MOSFET has three terminals, a source where electrons enter, a gate which controls how many electrons pass through the channel and a drain, where they exit the transistor. The gate is insulated from the channel, presenting a high electrical driving impedance. By reducing the channel lengths and widths to a few μm or less, output impedances and signal transit times have become very small. In fact, it has come to the point where time delays resulting from the resistances and parasitic capacitances of the interconnects between transistors are the limiting factors in circuit speed.

A new arrangement called CMOS (Complementary Metal Oxide Semiconductor) technology was introduced in 1963. Here n-channel MOSFET's are paired with p-channel (hole-conducting) MOSFET's in a way that essentially no electrical power is dissipated except when they are switching on-off states. Power consumption is reduced and immunity to electrical noise is much higher than for NMOS circuitry. For many years, CMOS mainly played a role in wrist watches, clocks, timers and other low speed counting or calculating devices. Partly because of the complexity of parasitic effects, it

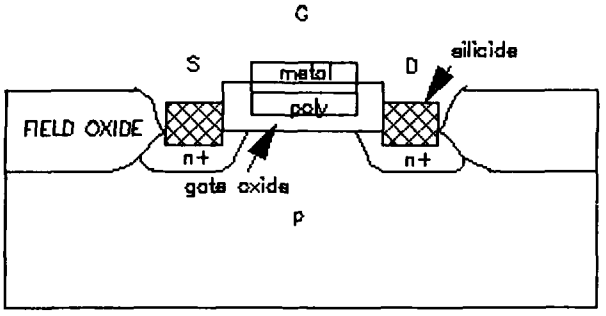


Fig. 1 n-channel MOSFET

was never thought to be amenable to very large scale integration into high speed digital circuitry. This has changed rapidly in the last five years and CMOS is now heralded as the dominant digital technology of the future. In order to bring CMOS into the high speed arena, many added process steps have been necessary and the material complexity of the circuit has been increased.

The volume of literature on silicon technology and related research is enormous, easily counted in thousands of publications per year. Many volumes on topical symposia have been published in the last ten years. More recent volumes of this sort are referenced first here[1-6]. The other references appended to this article are mainly either highly significant review papers or landmark individual publications. The reader's curiosity about the pioneer invention or publication in a given process or device area can be satisfied by referring to the detailed review articles. Three in particular deeply contrast the technology to the LSI era[7], to the VLSI period[8] and the approach to ULSI (Ultra Large Scale Integration, 2M-64M components per chip)[9].

SILICON

Either in the final circuit or at some time in the fabrication sequence, this basic element is seen in three forms, single crystal, polycrystalline (hereafter called polysilicon) of widely varying grain size or amorphous. This versatile material tetrahedrally bonds in the diamond crystal structure with an interatomic distance of 2.34 A. The crystal has a density of 2.33 g/cm^3, corresponding to an atomic density of approximately 5 X 10^{22}/cm^3. Theoretically pure silicon at room temperature has intrinsic carrier concentrations of holes and electrons of 1.38 X 10^{10}/cm^3.

Table 1 IN SI TECHNOLOGY

THE GOOD GUYS	THE BAD GUYS
O_I	O_I
B	ALKALI IONS
P	Na^+, K^+, Li^+
As	TRANSITION METALS

The openness of the covalently bonded structure allows a much greater variety of point defects than is customary in metals. The most obvious are Group III atoms, of which boron is the most useful, or Group V atoms, where phosphorus and arsenic are commonly used, substituting for silicon atoms to give p-type and n-type extrinsic conductivity, respectively. The useful doping range is roughly 10^{15}-10^{21} atoms/cm^3.

Vacancies, self-interstitials, and both substitutional and interstitial solute atoms play a role in diffusion phenomena in silicon. From the early '60's to the mid-'70's, diffusion from a glassy source through holes in a thermally grown SiO_2 mask was the principal method of impurity doping. As more and more unusual diffusion and precipitation phenomena are discovered from the circuit fabrication technology, more complex thoretical interpretation is needed[10,11]. A completely satisfactory result has yet to be fulfilled[12].

Several solute atoms added either intentionally or unintentionally to silicon are separated in Table 1 into beneficial or harmful. Unintentional impurities may come from the crystal growing process, such as interstitial oxygen. Other unintentional and unwanted impurities can arrive from surface contamination or from furnace heaters or residues in later processing. Examples are the alkali ions and transition metals, such as Cu, Ni and Fe.

Oxygen is shown as both good and bad. It is peculiar to CZ crystal and usually ranges from 10-20 ppm (approximately 0.5-1 X 10^{18}/cm^3). Interstitial oxygen in silicon, O_I, is accurately measurable from an infrared absorption peak. It was recognized early that CZ crystal resisted plastic deformation during and after high temperature processing much more than float-zone silicon, which is low in O_I. It is now clear that O_I strengthens silicon wafers greatly by pinning dislocation movement[13,14,18].

The greatest difficulties arose when the movement of oxygen to substitutional sites and its precipitation to SiO_2 (thought to be cristobalite) occurred uncontrollably in the active regions of the circuit. This difficulty was turned to a benefit in the early '80's by closely controlling the amount of O_I in the crystal and localizing the precipitation deep in the wafer, leaving what is called a denuded zone at the surface where the circuitry is built. This is called intrinsic gettering[15].

Gettering is a procedure dating back to the earliest days of silicon technology, whereby transistor yields are improved by certain poorly understood process steps. Early ones were wafer backside damage, phosphorus backside diffusion[16], HCl addition[17] and polysilicon addition to the backside[18]. It was thought that transition metal atoms and other undesirable impurities precipitate deep in the wafer, thus well away from damaging the active circuit elements. If the oxygen precipitation IG is overdone, mechanical strength under thermal stress during processing decreases radically[18]. A defect density limit in the bulk is reported as 5 X $10^9/cm^3$ to prevent large wafer warpage[14]. An excellent review of the evolution of the careful control of O_I in the CZ crystal has been given recently[19].

Silicon is sometimes thought of in solid-state electronics as the analogue of steel in mechanical applica-tions. Indeed, in the last 25 years, the diameter of crystal and therefore the virgin wafer has increased from one inch to the current fab plant size of 150 mm (nominally 6 inches). Crystals are being grown now in limited quantity at 200 and 250 mm (8 and 10 inch) diameter. In the fall of 1987, it was announced that Kayex Corp., a US maker of crystal growing furnaces, was shipping to Japan a unit capable of holding a 150 kg (330 lb.) charge and intended to produce 6, 8 and 10 inch diameter crystals in lengths up to ten feet. True to the steel analogy, these are being called ingots.

These growers are usually heated by electrical resistance and complex sensor arrangements with computer control leads to accurate and constant diameter. Both crystal and crucible are counter-rotated and magnetic fields are now being used to control convection in the melt[19]. Effort is expended for low temperature gradients in the pulled crystal. A dislocation-free seed is used and these large crystals are essentially dislocation free. Ironically, the formation of dislocations and stacking faults occur later in the circuit processing, but with ever greater control in recent years.

EPITAXIAL GROWTH

Silicon crystals are usually grown in either the (111) or (100) orientation. The (111) was favored in the '60's, but fell from favor when epitaxial growth became important in the bipolar transistor design. A buried arsenic-doped collector was often covered with a more lightly doped epitaxial body, allowing a nice compromise of high collector avalanche breakdown voltage and low series resistance.

A smaller growth rate, a tendency to pyramidal defects in the layer and to a "washout" or smearing of the pattern was characteristic of (111) wafer. The latter two problems could be corrected with a misorientation of a few degrees. However, a change to (100) led to improvements in all three respects and (100) gradually became the more common choice. A critical review of the effects of orientation shows that there are also some disadvantages in the use of (100) wafers[18].

Although epitaxial layers fell from design favor as the NMOS technology developed, a resurgence of their use occurred in the early '80's. Their use greatly reduces the nagging problem of parasitic latchup in CMOS circuits and increases the hold time of DRAM's. Part of the improvement versus epi-structures of the '60's results from use of IG[15].

The epi-layer is grown by chemical vapor deposition (CVD) with hydrogen reduction of silane, chlorosilanes or silicon tetrachloride at 800-1150°C. The single crystal surface substrate must be meticulously cleaned before and at temperature if a layer of high perfection is to be produced. Improved systems with complete automatic loading of wafer batches and careful particulate control have led to greatly reduced defects in the epi-layer.

ION IMPLANTATION

Since the mid-'70's, ion implantation has greatly displaced diffusion for the source doping of silicon. Diffusion furnaces are still used for driving in the implanted source, but even there Rapid Thermal Annealing (RTA), sometimes called Short Time Annealing (STA), is becoming the preferred step to follow ion implantation.

The advantages are several. Almost

any element can be implanted without solubility limitation. Control of the shallow layers needed in VLSI and ULSI is good. Freedom to tailor dopant profiles is greater than with diffusion. With modern equipment, the number of atoms in the dose can be metered quite accurately. However, a very sophisticated and costly machine is needed, wafer surface charging and overheating must be carefully avoided and damage to the crystal generally must be removed after the implant.

A troublesome effect known as channeling, or a deep penetration of ions in certain discrete crystal directions was originally reduced by a 7° misorientation of the beam with respect to <100>[20]. Eventually it was discovered that amorphous silicon was often formed at higher current high voltage implants. Furthermore, channeling was not a problem in the amorphous layer, as might be expected. Better depth and profile control of the implant is now accomplished by preamorphization with Si or Ge ions before the main doping[9,21]. Some angular misorientation is often still used, however.

Annealing after the ion implantation is usually needed for two reasons, to locate the dopant atoms on the substitutional sites so they will be active as donors and acceptors and to restore good crystal perfection and thus satisfactory carrier mobility. The proper regrowth of the preamorphized layer is especially critical and is termed solid phase epitaxial growth (SPE). However, very little dopant vertical or lateral diffusion is desired in the shallow layers now being designed. This challenge is especially critical in that high temperature effects are cumulative because of the many process steps.

This limitation on cumulative high temperature effects is quantified as "thermal budget". STA or RTA are used regularly for these anneals[18,22-23], which might typically take a wafer to 1100°C for 10 seconds and cool it rapidly back to room temperature. Because of its versatility for other process steps than post-implant annealing, RTA is often being called RTP (Rapid Thermal Processing)[24].

Several manufacturers are producing equipment with power sources ranging from lasers to intense tungsten-halogen lamps. An entire 1µm CMOS technology has been developed with a thermal budget of only 10 seconds at 1050°C[25]. These processes also favor the use of (100) wafers where dislocation networks spread less and fewer residual defects are left after annealing than in (111)[18].

POLYSILICON

Long before polysilicon appeared as a gate contact and interconnect in the NMOS technology, it was the raw material for charging the CZ crystal grower. Polysilicon or "poly" as it is often called, has been produced from the hydrogen reduction of silane or chlorosilanes which have been extremely purified.

A similar CVD process has been used since the early '70's to produce the polysilicon elements in MOS devices as in Fig. 1. As with most CVD, conformality is good, a requirement for virtually all films in VLSI. The use of low pressure (LPCVD) systems has lowered the deposition temperature drastically and improved the quality and uniformity of the film. Using silane and pressures of a few tenths Torr, deposition can be carried out at 550-650°C[26].

Polysilicon replaced aluminum as a contact and interconnect material mainly for two reasons. First, its higher melting point allows the metallization to act as a self-aligning mask for later high temperature treatments such as source and drain diffusion. Second, in VLSI interconnects must cross over. This multilevel situation usually requires the formation of oxide layers between interconnects at temperatures above the melting point of aluminum. The one great disadvantage is the high resistivity of polysilicon, about 400 µohm-cm when highly doped with phosphorus, as contrasted with 3µohm-cm in the displaced aluminum. The solution has been to form silicides of transition or refractory metals with resistivities in the range 10-50 µohm-cm augmented with refractory metals, such as tungsten in the 5 µohm-cm range, while still maintaining a high melting point. Other papers in this Congress discuss silicides and CVD processing in detail.

There has been a continuing interest in understanding the electrical conduction mechanism in polysilicon films and operative transistors have been made with them[27]. The incorporation of hydrogen into the film with the use of plasma enhancement (PECVD) has improved the performance of these devices. At a certain hydrogen level, the silicon becomes amorphous and has been commercially important for photovoltaic arrays. It is thought that hydrogen stabilizes the amorphous structure by attachment to dangling silicon bonds.

ISOLATION TECHNOLOGY

Isolation of one circuit element from another has been a challenge since the beginning of circuit integration. In conventional siicon IC's, the active circuit extends a few micrometers into a single crystal chip of 300-600 μm thickness. Most of the difference is merely a convenient, mechanically supportive "handle" during manufacture. The ideal arrangement from a circuit performance standpoint would seem to be perhaps 10 μm thick elements completely surrounded with an insulator. There have been two principles for the increasingly complicated efforts to approach such ideal isolation. The first is speed and reliability of performance, the attainment of which must be balanced against cost in the commercial market. The second is the need of military organizations to have radiation-hardened circuitry which will function in intense radiation fields. Such behavior calls for a minimization of the actual circuit volume which is penetrated by the field. The economics of the second demand tolerates a much more costly process than that of the first. In fact, such circuits have been made many years for the military using a process where islands are prepared on a silicon wafer, they are isolated completely with an oxide, a "handle" of polysilicon or other material is put on top and the single crystal wafer is carefully removed to the isolation level. In addition to cost, there are several control difficulties with this approach.

On the commercial side, partial isolation was advanced in the early '70's with a process called LOCOS (Local Oxidation of Silicon)[28]. By the use of silicon nitride, Si_3N_4, which serves as an excellent mask for the oxidation of silicon, a field oxide of as much as 2 μm in depth, as shown in Fig. 1, is produced. When the nitride alone was used, many dislocations were formed during the oxidation under the edge of the nitride areas. These were found to be harmful to the circuits and a thin pad (20-40 nm) of LPCVD SiO_2 was introduced between the nitride and silicon to buffer the dislocation inducing stresses.

The thermal expansion stresses and thus the defects were reduced thereby, but a change in shape of the field oxide resulted in what is called the "bird's beak". The field oxide to gate oxide transition in MOS increases to about 1μm or more and the planarity of the surface is reduced. This form of LOCOS became standard in the industry with some modifications. By 1982, the advent of 1-2 μm VLSI found the lost packing area intolerable and efforts had intensified in three directions:

1) Significant efforts to eliminate bird's beak by significant process changes to LOCOS such as SILO (Sealed Interface Local Oxidation) and SWAMI (Side Wall Masked Isolation)[29-32].

2) Apply new RIE (Reactive Ion Etching) and CVD techniques to deep trenching and backfill processes. Some defect generation problems similar to LOCOS are being encountered however[33,34].

3) Work on SOI (Silicon-on-Insulator) technology which now has led to a serious comparison of four approaches, SIMOX (Separation by IMplanted Oxide), SIMNI (Separation by IMplanted Nitride), ZMR (Zone Melted Recrystallization) and FIPOS (Full Isolation by Porous Oxidized Silicon)[35-38].

These four systems for obtaining full isolation are shown schematically in Fig. 2. They have all been worked on for several years. Ideas on SIMOX began in the early '70's and actual work began late in that decade.

SIMOX and SIMNI are similar in approach. A continuous insulating buried layer of silicon oxide or silicon nitride is formed by high energy implantation of oxygen or nitrogen ions. This layer is formed at sufficient depth below the surface to leave a layer of single crystal silicon in which circuits may be fabricated. The practicality of this process was hampered for a long time by the lack of a high voltage implanter which would deliver a reasonably large ion current. A dose of approximately $10^{18}/cm^2$ at 150-200 keV is required. The recent development of machines which will deliver tens of milliamps has reduced the time for the implant to reasonable levels. There are two main differences between SIMOX and SIMNI. First, because of its high diffusivity, excess oxygen will increase the thickness of SiO_2, whereas nitrogen, with low diffusivity, can accumulate in excess of that needed for Si_3N_4. Second, the nitride layer recrystallizes during post-implant annealing, but the oxide remains amorphous[37]. The wafer is heated to 400°C or higher during implant in order to maintain the crystallinity of the top layer of silicon[39].

In one form of ZMR, a polysilicon layer, typically 0.5 μm thick, is deposited by LPCVD on a thermal SiO_2 layer, 0.5-2 μm thick, which has been formed

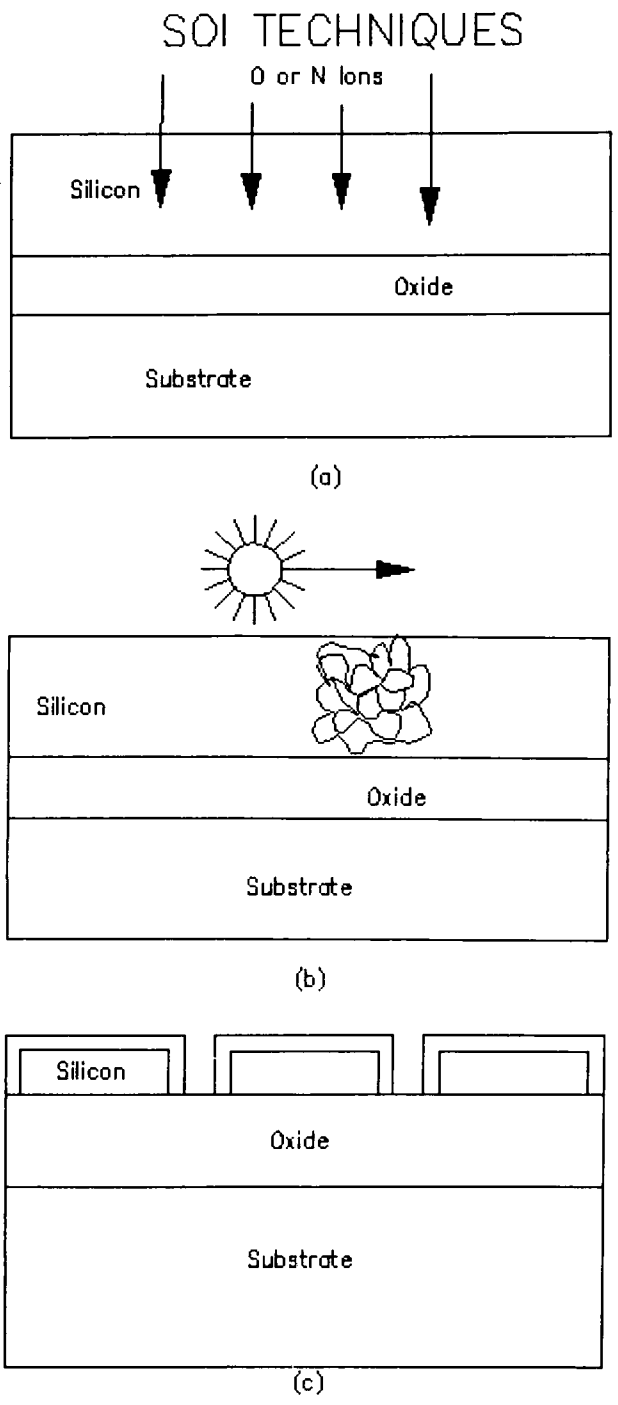


Fig. 2 Four Principal SOI Techniques (a) SIMOX & SIMNI, (b) ZMR and (c) FIPOS. After Weaver[38], modified.

on a single crystal wafer. An encapsulating layer of CVD SiO_2, 2μm thick, is placed over the polysilicon. A graphite strip heater system passes over the sandwich forming a molten layer. A seed can be used, but it is not common. Unseeded films show a (100)<100> texture and contain low angle boundaries with 1° of misorientation. Working CMOS LSI circuits have been made from these films[40]. Other workers in this general area have used SPE (Solid Phase Epitaxy) growing single crystal films on sapphire substrates (SOS)[41].

FIPOS depends basically on preferential electrolytic anodization of layers of p-type silicon to form a porous structure surrounding islands of n-type crystal. During subsequent oxidation, the porous regions are oxidized rapidly and preferentially. There are several variations in the methods of initiating and delineating the porous layer. CMOS circuits of good performance have been made in the isolated single crystal volumes s wide as 325 μm[42].

Strong government contract support for these programs have led to critical analysis, particularly comparison of their advantages and disadvantages for substrates in CMOS fabrication[38,43].

Yet another entry in the progression to full isolation is a field assisted process whereby application of a moderate voltage between two oxidized silicon wafers converts them into a $Si/SiO_2/Si$ sandwich. Examination of 3 inch wafers so bonded showed them to be free of dislocations or stacking faults[44].

REFERENCES

1. VLSI Science and Technology/1982, Ed. by C.J. Dell'Oca and W.M. Bullis, Electrochemical Society Proceedings, Vol. 82-7(1982).

2. VLSI Science and Technology/1984, Ed. by K.E. Bean and G.A. Rozgonyi, ECS Proceedings, Vol. 84-7(1984).

3. VLSI Science and Technology/1985, Ed. by W.M. Bullis and S. Broydo, ECS Proceedings, Vol. 85-5(1985).

4. Semiconductor Silicon 1986, Ed. by H.R. Huff, T. Abe, and B. Kolbesen, ECS Proc. Vol. 86-4(1986).

5. Silicon Nitride and Silicon Dioxide Thin Insulating Films, Ed. by V.J. Kapoor and K.T. Hankins, ECS Proc. Vol. 87-10 (1987).

6. ULSI Science and Technology/1987, Ed. by S. Broydo and C.M. Osburn, ECS Proc. Vol. 87-11(1987).

7. B.E. Deal and J.M. Early, J. ECS 126, 20C-32C(1979).

8. A. Reisman, Semiconductor Technology, ECS Proc. Vol. 82-5, 221-249(1982).

9. C.M. Osburn and A. Reisman, J. Elec. Matr. 16, 223-243(1987).

10. T.Y. Tan and U. Gosele, Ref. 2, 151-175(1984).

11. S.M. Hu, Ref. 3, 465-506(1985).

12. U. Gosele, Ref. 4, 541-555(1986).

13. F. Shimura, Ref. 1, 17-32(1982).

14. S. Kishino, Ref. 3, 399-418(1985).

15. C. W. Pearce and G.A. Rozgonyi, Ref. 1, 53-59(1982).

16. J.M. Andrews, P.A. Heimann and R.A. Kushner, Ref. 1, 68-75(1982).

17. T. Hirao and S. Maegawa, Ref. 4, 927-938(1986).

18. S.M. Hu, Ref. 4, 722-750(1986).

19. J.W. Moody, Ref. 4, 100-116(1986).

20. M.I. Current and K.A. Pickar, Semiconductor Technology, ECS Proc. Vol. 82-5, 65-88(1982).

21. S. Solmi, S. Guimaraes, E. Landi and P. Negrini, Ref. 4, 583-593(1986).

22. T.O. Sedgwick, Ref. 1, 130-146(1982)

23.T.E. Seidel, R. Knoell, F.A. Stevie, G. Poli and B. Schwartz, Ref. 2, 201-210(1984).

24. S.R. Wilson, Ref. 4, 621-641(1986).

25. D. Sharma, S. Goodwin-Johansson, D-S. Wen, C.K. Kim and C.M. Osburn, Ref. 6, 49-63(1987).

26. B.S. Meyerson, Ref. 4, 223-234(1986).

27. T.I. Kamins, Ref. 4, 235-249(1986).

28. J.A. Appels, E. Kooi, M.M. Paffen, J.J.H. Schatorje and W.H.C.G. Verkylen, Phillips Research Reports 25, 118-132 (1970).

29. W.G. Oldham, IEDM Digest, 216-219 (1982).

30. J. Hui, T.Y. Chiu, S. Wong and W.G. Oldham, IEDM Digest, 220-223(1982).

31. K.Y. Chiu, R. Fang, J. Lin, J.L. Moll, C. Lage, S. Angelos and R. Tillman, IEDM Digest, 224-227(1982).

32. C.W. Teng, W.R. Hunter and R.A. Haken, Ref. 4, 515-526(1986).

33. R.D.Rung, H. Momose, and Y. Nagakubo, IEDM Digest, 237-240(1982).

34. S.Y. Chiang, K.M. Cham, D.W. Wenocur, A. Hui and R.D. Rung, Ref. 1, 339-346 (1982).

35. R.J. Chater, J.A. Kilner, P.L.F. Hemment, K.J. Reeson and R.F. Peart, Ref. 4, 652-666(1986).

36. A.E. White, K.T. Short, L.N. Pfeiffer, and K.W. West, Ref. 6, 723-731(1987).

37. H.J. Stein, Ref. 5, 429-447(1987).

38. SOI Technology, Ed. by H.T. Weaver, IEEE Circuits and Devices, Pt. I, July 1987, Pt. II, November 1987.

39. H.W. Lam, Ref. 38, 6(July 1987).

40. B.Y. Tsaur, Ref. 38, 12(July 1987).

41. P.K. Vasudev, Ref. 38, 17(July 1987).

42. S.S. Tsao, Ref. 38, 3(November 1987).

43. B.Y. Mao, R. Sundaresan, C.E.D. Chen, M. Matloubian and G. Pollack, IEEE Trans. ED, 35, 629-633(1988).

44. R.C. Frye, J.E. Griffith and Y.H. Wong, J. ECS 133, 1673-1677(1986).

NEW DIRECTIONS FOR SEMICONDUCTOR DEVICE RELIABILITY

Jay W. Lathrop
Center for Semiconductor Device Reliability Research
Clemson University
Clemson, South Carolina 29634-0915, USA

ABSTRACT

Advances in integrated circuit fabrication technology over the past two decades have increased chip complexity to 10-million transistors per chip, while simultaneously decreasing failure rates by orders of magnitude and increasing yield. This tremendous technical achievement, which has come about as a result of greatly reducing manufacturing defects, has required rethinking traditional approaches to reliability assurance for two reasons. As failure rates due to defects approach undetectable levels, it is no longer practical to attempt to measure reliability in an absolute sense. In addition, as dimensions shrink there is concern that at some point wearout mechanisms will begin to dominate over defect related failure mechanisms. The paper illustrates how computer aided methods can be used to control and manage reliability using as examples three types of integrated circuit failure mechanisms: charge injection, electromigration, and electrostatic discharge overstress.

RELIABILITY REFERS to the probability that the electrical characteristics of a device, which initially lay within specified limits, will remain within these limits for some specified period of time. Much effort is expended each year by industry in support better component reliability. Technological factors, however, are rapidly changing the nature of reliability problems in the industry. This paper reviews historical trends in integrated circuit reliability, showing how reliability assurance techniques for future devices will need to be different. The status of several specific failure mechanisms will be discussed to show how computer models can be combined with a knowledge of material properties to improve the reliability of future devices.

CLASSIFICATION OF FAILURE MECHANISMS

The classical "bathtub" curve of Fig. 1 indicates the three classifications of failure mechanisms: wearout, defect accelerated, and overstress.

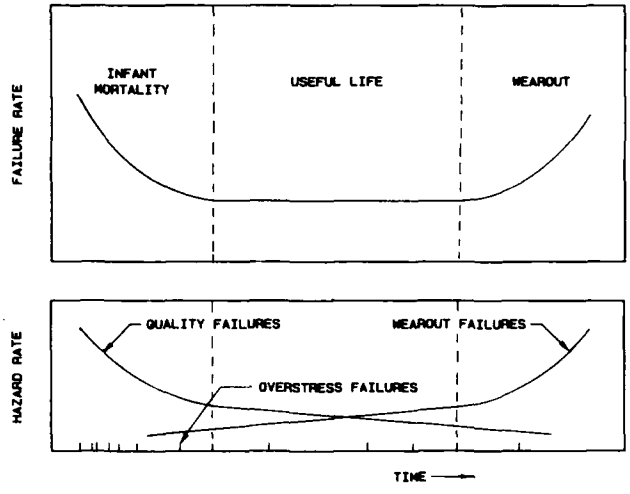


Fig. 1. Classical failure rate vs time curve

WEAROUT - Wearout refers to degradation of the main population of non-defective devices and ideally will occur a considerable time after manufacture. Examples of environmental wearout are corrosion, surface ionic inversion, and diffusion, while examples of operational wearout are electromigration and charge injection.

DEFECT ACCELERATED WEAROUT - Defective devices, i.e. those involving freak distributions of one or more physical parameters, will also be affected by environmental and operational wearout mechanisms. However, because of the presence of flaws, either introduced during manufacture or present in the starting material, defective devices degrade faster than those of the main population. Defect accelerated wearout leads to the "infant mortalities" of Fig. 1.

OVERSTRESS - The third category of failure mechanism is that of overstress. The term can include mechanical or thermal overstress, but most frequently is used to designate electrical overstress in the form of electrostatic discharge (ESD) or electrical transients. Exposure to ESD may occur at any time during life, but is more likely soon after manufacture when components are handled, tested, inserted in boards, etc. ESD is shown schematically in the bathtub curve of Fig. 1 as a series of discrete spikes.

IMPACT OF TECHNOLOGICAL CHANGE

The reliability of semiconductor devices has always been of concern, but the evolution of technology has greatly changed the nature of this concern. Traditionally, IC reliability has been concerned with infant mortality. Wearout was in general not considered a problem because, by extrapolating accelerated test results to use conditions, it could be shown that devices would almost certainly become obsolete long before the main distribution would be affected.

Fig. 2 illustrates the exponential increase in chip complexity that has occurred as a result of technological developments in the integrated circuit industry. Single chip IC devices, which currently contain as many as 10-million transistors in 100+ lead packages, by the mid to late 1990's can be expected to contain as many as 1-billion transistors in 400+ pin packages. As can be seen from this figure, average failure rates over the past two decades have shown a decrease as spectacular as the increase in complexity.

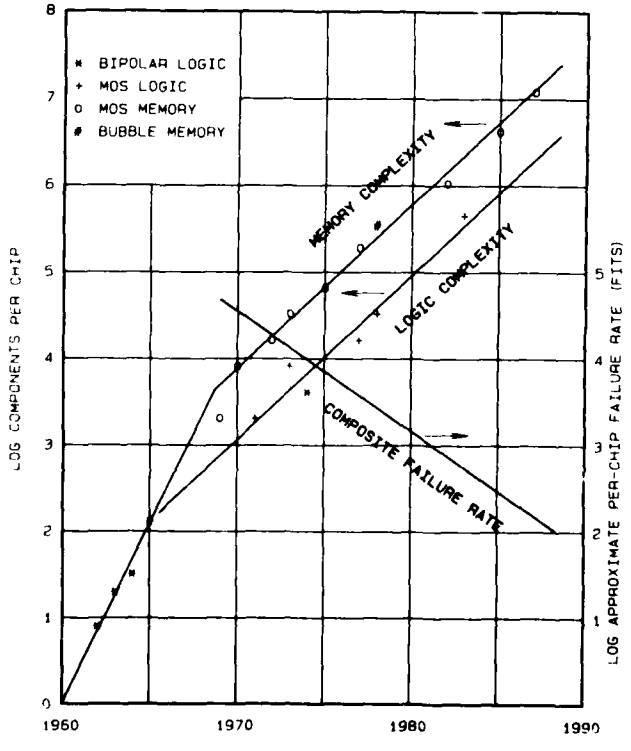


Fig. 2. IC complexity and failure rate vs time

Equally amazing was an accompanying increase in yield. The historical correlation between failure rate and yield leads to the premise that yield shields reliability. That is, that because low yield devices cannot be produced economically low reliability devices will not be a problem. This type of thinking, however, overlooks both the increased reliability expectations that occur with time as IC applications become more numerous and more critical, and the possibility that as device dimensions shrink, wearout processes may no longer be ignored.

COMPUTER SIMULATION

The goal of reliability research is to be able to predict the probability of device failure under use conditions and to take corrective action if the predicted probability is unsatisfactory. Infant mortality failure rates are now so low as to be almost undetectable, even under accelerated test conditions. Under these conditions, the only practical alternative is computer simulation of relative failure rates, permitting design and process comparisons. Simulation can also determine conditions under which wearout failure mechanisms become limiting.

Computer simulation involves the following sequence of events: 1) characterization, 2) mathematical modeling, 3) incorporation of the model in appropriate computer aided design (CAD) and manufacturing (CAM) tools, and 4) use of these tools to optimize design. Understanding the physical principles of various failure mechanisms will be the key to successful implementation of this sequence. It is important to note that implementation will make allow reliability to be traded off against cost and performance. Three failure mechanisms of particular concern to very large scale integrated (VLSI) devices, electromigration (EM), charge injection (CI), and electrostatic discharge (ESD), will now be discussed with particular emphasis on computer analysis and simulation.

ELECTROMIGRATION

Electromigration is a clear example of a wearout mechanism. It originates as a result of momentum transfer between current carrying electrons and metal atoms of the interconnect metallization. This momentum transfer results in an atomic flux, J, which is proportional to the dot product of mobility and driving force [1]. Mass transport can take place through either point defects, such as vacancies or interstitials, or gross defects, such as grain boundaries or surfaces. At normal device operating temperatures, which are less than half the melting temperature of the metal interconnects, mass transport will be primarily along the grain boundaries. It has been found that addition of a few atomic percent of Cu into the Al metallization reduced the EM flux. To be

effective, however, the Cu must go into the grain boundaries rather than into the grains themselves, since Cu within the grains merely increases the resistance. "Stuffing" the grain boundaries with Cu is achieved by first depositing Cu on the metal film and then, through a series of heat treatments, drive the Cu to the grain boundaries where it is formed into the Al_2Cu alloy.

Another approach to the reduction of voiding is to avoid the propagation of EM cracks completely through the conduction path by using multilayer films. If one of the layers has a higher resistance, so that its current density is less, and/or if it has a higher atomic number so that momentum transfer is less efficient, then current could still flow, even if the other layer were to open up. While such a structure should be very effective against voids, shorts could still be expected.

Harrison [2] has proposed a statistical EM model based on first principles which includes backflow due to self diffusion and pressure. This model approximates the metallurgical state of the line using Monte-Carlo statistical methods. A log-normal grain size distribution is assumed, which is characterized by a median value and a standard deviation. The modeled line is partitioned into sections whose length is equal to the average grain size. Boundaries between segments are called nodes and the mass flux divergence of each node calculated. By combining the effect of momentum transfer induced mass transport along the line with backflow effects [3], it is possible to calculate the loss of mass at each node. The increase in local temperature and degradation of heat transfer due to current crowding as mass is lost are also included in the model. Finally, the increase in resistance of the line due to the reduction in cross sectional area is calculated to allow a direct comparison with experiment.

The model has been used [4] to examine the effect of random defects along a line on the change in resistance of the line with time under dc current stress. Simulation examples for two lines, with and without the introduction of defects, are shown in Fig. 3. It can be seen that line B, which showed failure at 240 hours without defects, failed after only 15 hours when defects were introduced in the form of notches at 1/4, 1/2, and 3/4 of the lines length (B'). On the other hand, line A, which showed a consistent, but non-failure increase in resistance without defects, changed by a negligible amount when these same defects were introduced (A'). Thus the location of defects relative to the random grain boundaries appears significant and the next step will require probabalistic considerations.

Once a model has been developed which can account for EM under all conditions of current, including pulsed current, and in leads of different geometry (steps, vias, contacts, etc.), it can be included in available computer aided design tools to calculate the reliability

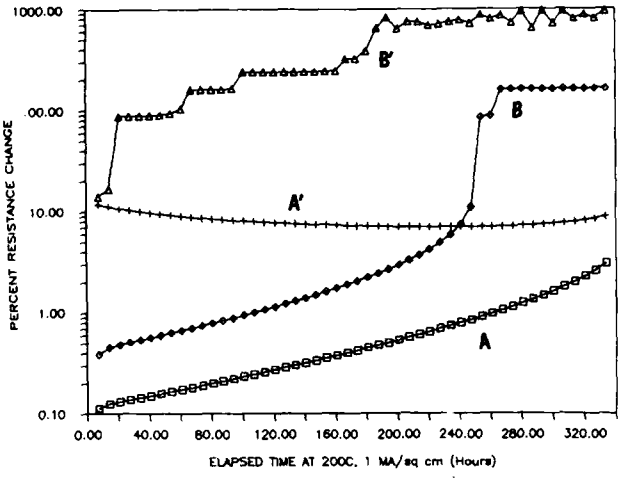


Fig. 3. Simulated resistance change with time for two lines with (A',B') and without (A, B) the introduction of equal defects.

contribution of EM. Preliminary work by Frost and Poole [5] based on SPICE simulation has demonstrated the feasibility of reliability simulation and at the same time pointed out the need for better models and more cost effective approaches. Fig. 4 shows the use of their program RELIANT to simulate EM wearout in a simple CMOS inverter. Although this example is clearly trivial the rising wearout portion of the failure rate curve has been simulated. As more information is received regarding the experimental behavior of EM test vehicles, the first order tools can be refined and limiting constraints eliminated. Analysis capable of locating critical areas [6] can then be followed by more thorough design optimization.

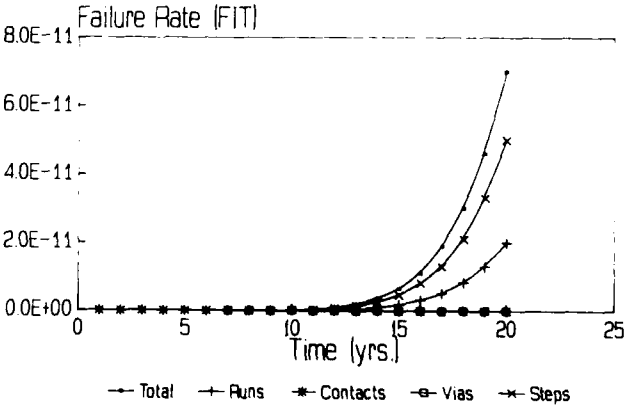


Fig. 4. Failure rate simulation for simple CMOS inverter using RELIANT program.

CHARGE INJECTION

As integrated circuit feature size is reduced, a corresponding reduction in thickness of the oxide dielectric layers must also take place. Sufficiently thin oxide layers will permit the passage of electrons. At thicknesses

greater than 10 nm current is conducted primarily by hot electrons, but below about 10 nm, oxide films are so thin that appreciable numbers of thermal electrons (or holes) are able to transit by means of Fowler-Nordheim tunneling. Passage of electrons through an oxide can change its insulating and dielectric properties and consequently the characteristics of devices which utilize these properties.

Hot electrons may be injected into the oxide from the channel, from the avalanche plasma near the drain region and, as a result of the existence of a high transverse field, from thermally generated carriers in the bulk. These hot electrons will charge the oxide, increasing the subthreshold leakage and changing the threshold voltage [7]. However, the contribucion of hot electrons to device degradation is expected to lessen as devices shrink for two reasons: 1) geometrical changes such as buried channels and lightly doped drains (LDD) have been incorporated in devices to minimize the number of hot electrons injected in the oxide, and 2) both the subthreshold leakage and the threshold voltage shift are inversely proportional to the capacitance per unit area, C_o, which increases as the oxide layer gets thinner. For these reasons, Threshold shifts due to oxide charging probably will not represent a major reliability problem in future VLSI devices, except for non-volatile memories and certain critical sense amplifier circuits.

This does not mean, however. that CI does not represent a potential reliability hazard. There is evidence that breakdown in thin oxides depends on the oxide's current and voltage history. This is the so-called time-dependent-dielectric-breakdown (TDDB) or charge-to-breakdown (QTB) effect. TDDB can be studied either by the application of a constant voltage stress and observing I vs t or by the application of a constant current stress and observing V vs t. Another technique often referred to (inappropriately) as time-zero dielectric breakdown (TZDB) involves the use of a voltage ramp.

Thin oxide capacitors subjected to constant voltage stressing have been shown to breakdown after hole fluence has reached a critical value, which depends on the electric field in the oxide E_{ox} [8]. In this case, the time-to-breakdown, t_{bd}, is an exponential function of $1/E_{ox}$ over many orders of magnitude as shown in Fig. 5 [9].

Localized high field regions undoubtably play a major role in TDDB since they alter the electric field at the cathode reducing t_{bd}. These occur because of such micro irregularities as interface roughness, nonuniform trap distribution, or dopant impurity segregation at grain boundaries. Thus the susceptibility of oxides to breakdown will depend on both the electrical history of the oxide and the statistical distribution of process related defects. This would imply that some form of continuous process monitoring coupled with the software evaluation of specific designs will be necessary.

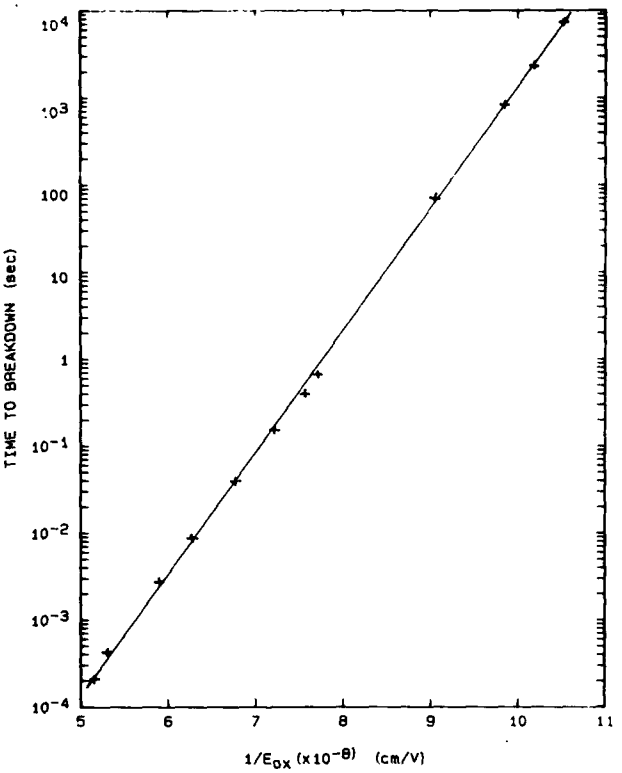


Fig. 5. Time to breakdown vs $1/E_{ox}$ for a 79 Å oxide capacitor. (After Lee, Chen and Hu)

Before a program like Frost and Poole's RELIANT can be used to compare designs on their susceptibility to CI effects it will be necessary to clearly identify the key variables and relate them to device operation. Promising candidates appear to be fluence and field. Once the variables have been identified a model needs to be developed which permits extrapolation to use conditions. It is relatively easy to obtain a measure of oxide robustness under accelerated conditions using capacitor test structures, but it is not yet clear how this can be translated into the mean time to failure of transistor gate oxides under use conditions.

ELECTROSTATIC DISCHARGE

Protecting semiconductor devices from ESD damage represents an important reliability effort. One aspect involves avoiding exposure to ESD pulses, activity which encompasses a vast area of physics relating to triboelectricity. In this paper, however, discussion will be restricted to the interaction of ESD pulses with semiconductor devices rather than generation and avoidance methods. This "after the fact" approach to ESD involves the use of an on-chip protection circuit at each input and output pin which is capable of dissipating excess energy without damage. Exposure to ESD, which is of concern today, will be of even more concern in the future for two reasons: 1) greater chip

complexity, which may necessitate as many as 400 pins per chip, will require that large areas of the chip be devoted to protection circuits, and 2) reduced feature size/oxide thickness will make devices more susceptible to voltage overstress.

Historically designers of on-chip protection circuits have subjected test structures, incorporating different geometries and utilizing different process technologies, to injection pulses and then subsequently analyzed them for ESD damage. Empirical interpolation of this data has then been used to meet particular design objectives in terms of the injection pulses used. Because a certain amount of silicon area (volume) is required in which to dissipate the pulse energy, protective circuit design involves a tradeoff between area usage and the degree of protection desired.

In order to effectively quantify this tradeoff two things are needed: 1) a precisely defined and repeatable injection impulse testing system which can be related to ESD exposure in the field, and 2) a theoretical basis for ESD design methodology which can be integrated with existing CAD tools. Testing of ESD protection has historically involved use of both the human body model and the charged device model pulse shapes. Neither of these test methods is particularly reproducible nor is the relationship between the two well understood. Recently, square wave testing using the discharge of a charged transmission line has proven very reproducible [10] and there is indication that it can be related to both methods. This would open the way to use of a single test and eliminate much confusion in the industry.

Metal-oxide-silicon capacitors subjected to ESD impulses exhibit a number of phenomena which depend on the pulse shape and amplitude: hot electron trapping, oxide breakdown, filamentation, contact damage, and thermal runaway (second breakdown). The complete breakdown process, which occurs when sufficient voltage is applied, consists of two phases: electrical breakdown of the oxide, which occurs within a few nanoseconds, followed by thermal runaway, which requires roughly 10X as long. Electrical breakdown can be preceded by localized charge trapping, which alters field lines concentrating the current and thereby initiating thermal runaway. Thermal runaway, causes melting of the metal contact and silicon substrate and the subsequent ejection of molten material [11]. An examination of metal deposition patterns indicates that the molten material was ejected under the influence of the crossed electric and magnetic fields associated with the discharge current flow. This means that the time scale for the ballistic sequence must be of the same order as the electrical events. Furthermore it can be calculated that the average trajectory speed must be greater than 500 m/sec [12].

When pulses insufficient to cause complete breakdown are applied, filamentation can occur. Degradation rather than destruction can be observed in commercial protection circuits subjected to sub-breakdown impulses. Bridgwood [13] has shown that the degradation of commercial protection circuits can be modeled by parallel filaments shunting a reverse biased p-n junction, with the number of filaments dependent on the number of applied pulses. Fig. 6 illustrates how the slope of the IV characteristic, which represents the input impedance of the protection circuit, decreases with the number of applied pulses. Degradation of this sort is a potential source of latent failures. It has also been determined that protection circuit response depends critically on rise time. Information gained in the study of MOS capacitors can be used to develop circuit models which will allow protection circuits to be tailored to specific applications.

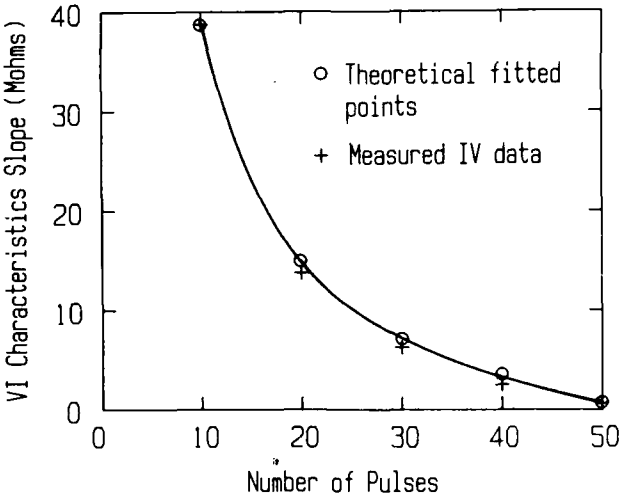


Fig. 6. Degradation of protection circuit input impedance due to sub-catastrophic ESD pulses.

Ultimately the objective of the ESD work is not so much to predict the probability of failure, as was the case for electromigration and charge injection, but rather to be able to use CAD methods to design protection circuits through a better understanding of the breakdown mechanism. Integrating protection circuit design with the total functional chip design should permit tradeoffs between cost, performance, and protection to be made.

CONCLUSIONS

Failure mechanisms in each of the three categories - defect accelerated, wearout, and overstress - will be of concern to future devices. Defect accelerated (infant mortality) failures will soon reach levels which are undetectable by conventional techniques. The challenge will then be to develop methods of improving reliability when the results can't be measured. This will need to involve software methods aimed at a relative assessment of designs and processes.

At some point in the evolution of integrated circuits wearout mechanisms will become limiting. Because this wearout a function of the main distribution and not of a freak distribution it is possible that the effect on yield will be minimal. Thus yield will not serve as a precursor of reliability difficulties. The challenge will be to develop computer aided design guidelines which will keep wearout (on an absolute basis) beyond obsolescence. The two areas of most wearout concern involve metallization and oxide dielectrics (electromigration and charge injection). Within a few years design procedures should be available for EM, but comparable CI design procedures appear to be considerably further off.

Concern with ESD will increase as devices evolve because there will be less desire to devote space or sacrifice performance to improve protection. In order to answer the question of when a device has enough protection it will be necessary to devise a CAD method of design that will permit protection to be traded off against size and performance. Implementation of this concept will depend on being able to relate geometry and layout to breakdown susceptibility.

ACKNOWLEDGEMENT

The author wishes to thank his colleagues at Clemson University, M.A. Bridgwood, D.J. Dumin, J.W. Harrison, Jr., and K.F. Poole for their assistance in reporting the status of the ESD, CI, EM, and CAD work respectively.

Support for the reliability research program at Clemson University is provided by the Semiconductor Research Corporation.

REFERENCES

1. H.B. Huntington, "Diffusion in Solids - Recent Development, Academic Press, New York (1975)
2. J.W. La hrop, et al, "Design Rules Hold Key to Future VLSI Reliability," Proc. 7th University/Government/Industry Microelectronic Symposium (UGIM), pp. 91-94, Rochester, NY (1987)
3. J.D. Venables and R.G. Lye, "A Statistical Model for Electromigration Induced Failure in Thin Film Conductors," Proc. 10th IEEE Int. Rel. Phys. Symp., p. 159, New York (1972)
4. J.W. Harrison, Clemson University, Clemson, SC 29634-0915, to be published
5. D.F. Frost and K.F. Poole, "A Method for Predicting VLSI Device Reliability Using Series Models for Failure Mechanisms," IEEE Trans. on Rel., R-36, pp. 234-242 (June 1987)
6. C.H. Stapper, "LSI Yield Modeling and Process Monitoring," IBM J. Res. Dev. p. 228 (May 1976)
7. T.H. Ning, C.M. Osburn, and H.N. Yu, "Effect of Electron Trapping on IGFET Characteristics," J. Electron. Mater., vol. 6, p.65 (1977)
8. C.-F. Chen, et al, "The Dielectric Reliability of Intrinsic Thin SiO_2 Films Thermally Grown on a Heavily Doped Si Substrate - Characterization and Modeling," IEEE Trans. Elect. Dev., ED-34, pp. 1540-1552 (July 1987)
9. J. Lee, I-C. Cheng and C. Hu, "Statistical Modeling of Silicon Dioxide Reliability," Proc. IRPS, pp. 131-138, Monterey, CA (April 1988)
10. M.A. Bridgwood, "Breakdown Mechanisms in MOS Capacitors Following Electrical Overstress," Proc. 8th Electrical Overstress/Electrostatic Discharge Symp., pp. 200-207, Minneapolis, MN (Sept. 1986)
11. D.G. Pierce, "Electro-thermomigration as an Electrical Overstress Failure Mechanism," Proc. 7th Electrical Overstress/Electrostatic Discharge Symp., pp. 67-76, Minneapolis, MN (Sept. 1985)
12. M.A. Bridgwood and R.H. Kelley, "Modeling the Effects of Narrow Impulsive Overstress on Capacitive Test Structures," Proc. 7th Electrical Overstress/Electrostatic Discharge Symp., pp. 84-91, Minneapolis, MN (Sept. 1985)
13. M.A. Bridgwood, Clemson University, Clemson, SC 29634-0915, to be published

KINETICS OF THE OXIDATION OF SILICON

R. H. Doremus
Materials Engineering Department
Rensselaer Polytechnic Institute
Troy, New York 12180-3590, USA

KINETICS OF THE OXIDATION OF SILICON

Robert H. Doremus
Materials Engineering Department
Rensselaer Polytechnic Institute
Troy, New York 12180-3590

Abstract

The oxidation rates of silicon and silicides show simple parabolic kinetics with close agreement with the permeation of molecular oxygen in vitreous silica for a variety of conditions. Deviations from parabolic kinetics, decreases in the permeation rate from that expected, and increases in silica density occur together. Strain in the silica layer can explain all these changes. There is no evidence that an interfacial chemical reaction is rate determining in the oxidation of silica. The initial oxidation rate is often lower, not higher, than expected.

SILICON DIOXIDE LAYERS GROWN on silicon are important elements in many electronic devices and integrated circuits. As circuit elements become smaller and insulating layers thinner, it is particularly desirable to control the rate of oxide growth for the thinnest layers. Lower temperatures of growth would also be advantageous. The purpose of this paper is to explore mechanisms of oxide growth so that kinetics of oxidation can be confidently extrapolated to lower temperatures than have been detailed experimentally.

This paper will start with a comparison of parabolic oxidation kinetics and permeation of molecular oxygen in bulk vitreous silica. Then deviations from parabolic kinetics are shown to have a strong correlation with a reduction in the parabolic growth coefficient. No data give oxidation rates appreciably faster than expected from the permeation data of Norton [1]; initial oxidation rates are not faster than expected. Attention will be focused on oxidation of silicon by oxygen, because much more data are available than for oxidation by water; nevertheless, similar considerations should apply for water oxidation. Interfacial reactions are shown to be probably too fast to control oxidation rates. A model for reduction of oxidation rates by strain in the oxide is consistent with experimental results.

DIFFUSION-CONTROLLED LAYER GROWTH

Many experiments show that silicon dioxide grows at the oxide-silicon interface [2]; thus oxidation is controlled by the diffusion of oxidant through the oxide film and its reaction with silicon at the interface. An equation for the diffusion-controlled growth of a layer of thickness L after time t is [3]:

$$L^2 = 2\, c_L D t/\rho \qquad (1)$$

in which c_L is the concentration of oxidant in the oxide at the oxide-gas interface, the concentration of oxidant is assumed to be zero at the oxide-silicon interface, ρ is the oxygen concentration in the oxide, and D is the constant diffusion coefficient of oxidant in the oxide. A parabolic growth coefficient B is often used; here

$$B = 2\, c_L\, D/\rho \qquad (2)$$

For many years it was thought that the oxidant diffused through silicon dioxide as oxygen ions [2,4]. However, oxygen, water and other gases dissolve

and diffuse molecularly through vitreous silica [1,5], and certain experiments on the oxidation of silicon show good agreement with respect to absolute value, temperature dependence, and pressure dependence of molecular oxygen and water diffusion, as calculated from eq. 1 [6]. Recent independent experiments of different kinds have provided strong support for molecular diffusion of oxygen and water [7-11].

The measurements of permeation of molecular oxygen by Norton [1] are compared with parabolic coefficients from oxidation data for silicon and silicides that fit Eq. 1 in Fig. 1. There is excellent agreement in absolute values and temperature dependence; both permeation and parabolic oxidation rates are also proportional to oxygen pressure.

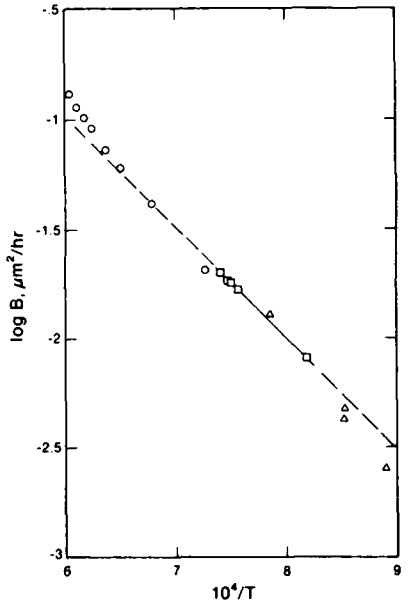


Fig. 1 - Logarithm of growth coefficients B from Eq. 1 for oxidation of silicon and silicides in one atm. of oxygen versus 1/T. ○ oxidation of silicon [4,16]; △ oxidation of silicides [2,17]; □ calculated from permeation data of Norton [1].

DEVIATIONS FROM PARABOLIC OXIDATION

At temperatures below 1100°C the kinetics of oxidation of silicon by both oxygen and water deviate from Eq. 1, the deviations becoming greater at lower temperatures. One way to examine these deviation is to plot log oxide thickness as a function of log time; a straight line results [12-14], and its slope gives a measure of deviations from parabolic growth, for which the slope is one-half, as shown in Table 1. There is a progressively greater increase in slope as the temperature decreases.

The following parabolic-linear-logarithmic (PLL?) equation fits all data on the oxidation of silicon and silicides by oxygen and water [12,14,15]:

$$L^2 + AL - AC\ \ell n(1+L/C) = Bt \qquad (3)$$

in which A and C are additional parameters. Coefficients B as calculated from fits to this equation [12], data from [4,18,19,20], are also given in Table 1. There is a close correlation between decreases in B and increases in the time exponent n for a power law fit from an equation such as

$$L = at^n \qquad (4)$$

in which a and n are parameters.

The density of the silicon dioxide film as calculated from its refractive index [21] increases as the oxidation temperature decreases, as also listed in Table I.

INTERFACIAL REACTIONS

In some phase transformations such as precipitation from liquid or solid solution there is evidence that both matrix diffusion and interfacial reactions influence transformation rates [14]. Thus it has been natural to explain deviations from parabolic oxidation kinetics as resulting from an interfacial reaction. A first order chemical reaction at the interface introduces a linear term such as the AL term in Eq. 3. Nevertheless there is no direct evidence that such a reaction is important in any oxidation kinetics. Bare metals and silicon react very rapidly with oxidants at temperatures above 500°C, and even at room temperature unoxidized silicon is quickly covered with a very thin protective oxide layer. There is no evidence for any metal or semiconductor that an interface reaction in rate determining [23]. Therefore it is unlikely that an interfacial reaction is slow enough to influence oxidation kinetics, and both the reductions in B and deviations from square root time dependence result from reduction of oxidant transport rates in the oxide.

RAPID INITIAL OXIDATION RATE

When oxidation data for silicon in oxygen are fit with certain equations such as that proposed by Deal and Grove, the initial rate is more rapid than expected from the fit. However, this initial rate is still slower than expected from a comparison with

Table 1 - Deviations from Square Root Dependence, Parabolic Coefficients, and Densities for Growth of Silica Films on Silicon in Oxygen

Temp. °C	B_N/B	Time Exponent n, eq.(4)	Density gm/cm^3
1200°	1.0	.50	2.198
1100°	1.04	.56	2.211
1000°	1.30	.62	2.224
900	2.09	.63	2.236
800	2.33	.66	2.254
700	6.85	.74	2.265

B_N - parabolic coefficient, eq. (1) calculated from data of Norton [1]

B - parabolic coefficient calculated for oxidation of silicon in oxygen

the data of Norton [1], as shown in Fig. 2. Thus it is not necessary to find mechanisms that increase the rate of oxidation, but only to explain why rates are lower than for unperturbed molecular diffusion and deviate from parabolic behavior.

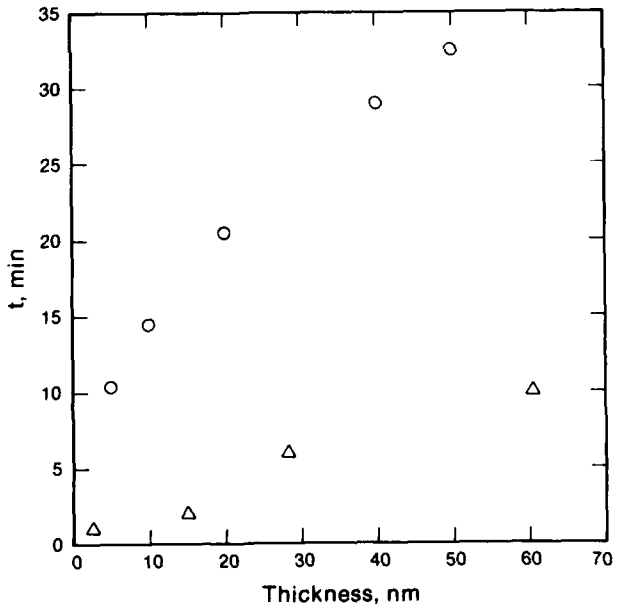


Fig. 2 - Thickness of silicon dioxide films growing on silicon in oxygen as a function of time at 800°C. ○ calculated for permeation data of Norton [1]; △ measured [19].

STRAIN

Strain in silica films on silicon reduces the rate of molecular transport in them [24], and also reduces the diffusion coefficients of molecules in bulk amorphous silica [27]. The refractive indeces of films grown above about 1100°C are close to that of amorphous silica, whereas at lower temperatures of film growth the refractive indeces (and densities) of films are greater than for bulk amorphous silica [20,26]. The changes in refractive index become greater at lower growth temperatures, corresponding to greater decreases in B values compared to those expected from Norton's data, as shown in Table 1. All these results strongly suggest that strain plays an important role in the deviations from parabolic kinetics.

A prediction from these considerations is that the refractive indeces of films for the data in Fig. 1, with permeations close to those measured for bulk silica, should be close to that of bulk silica. Increasing deviations from B values expected from these data should be accompanied by increasing deviations in the refractive indeces of the films involved. This prediction is perhaps the most direct way to test the importance of strain in the oxidation of silicon and silicides.

Tensile strain from quenching a glass causes higher ionic mobility, and hydrostatic compressive strain reduces the ionic mobility [27]; the mobility was proportional to the exponential of the change in specific volume of the glass. Thus it is assumed that the diffusion coefficient of molecular water or oxygen in amorphous silica depends exponentially upon the compressive strain ε:

$$D = D_L \exp(-k\varepsilon) \qquad (5)$$

in which D_L is the strain-free diffusion coefficient, and k can be related to an activation volume V*:

$$k = EV^*/RT \qquad (6)$$

where E is Young's modulus. It is assumed that the strain depends on the distance x in the film from the oxide-silicon interface as

$$\varepsilon = \varepsilon_o(1-x/L) \qquad (7)$$

where L is the total film thickness and ε_o the strain at the interface. From measurements of Cu_2O films on copper [28] it is assumed that ε_o depends on film thickness as

$$\varepsilon_o = 2q \quad 1 + \frac{\ell}{L+C} \qquad (7)$$

where q, ℓ, and C are parameters not dependent on film thickness. From Eq. 5-7 and equations for diffusion-controlled growth of the oxide film Eq. 3 can be derived with three fitting parameters:

$$A = \frac{2kq\ \ell}{1+kq}$$

$$B = \frac{2\ c_L D_L}{\rho(1+kq)} \qquad (8)$$

and C from Eq. 7. As described above Eq. 3 fits the oxidation data, and the parameters A and B have a reasonable temperature dependence for oxidation in both water [29] and oxygen [12]. Strain is therefore a likely explanation for the deviations from parabolic behavior and coefficients B.

Acknowledgement

This work was supported by NSF-MRC Contract No. DMR-85-10617.

References

1. F. J. Norton, Nature 191, 701 (1961).
2. P. J. Jorgensen, J. Appl. Phys. 37, 874 (1962).
3. R. H. Doremus, "Rates of Phase Transformations", Academic Press, Orlando, FL, 1985.
4. B. E. Deal and A. S. Grove, J. Appl. Phys. 36, 3770 (1965).
5. R. H. Doremus, "Glass Science", John Wiley, New York (1973).
6. R. H. Doremus, J. Phys. Chem. 80, 1773 (1976).
7. S. Rigo, F. Rochet, B. Agius and A. Straboni, J. Electrochem. Soc. 129, 867 (1982).
8. S. S. Cristy and J. B. Condon, ibid 128, 2170 (1981).
9. J. C. Mikklesen, F. L. Galeener and W. J. Mosby, Appl. Phys. Lett. 37, 712 (1980); 39, 403 601 (1981); J. Electron. Mater. 10, 631 (1981).
10. R. Pfeffer and M. Ohring, J. Appl. Phys. 52, 777 (1981).
11. B. Agius, S. Rigo, R. Rocket, M. Froment, C. Maillot, H. Roulet and H. Dufour, Adv. Phys. 35, 237 (1986).
12. R. H. Doremus and A. Szewczyk, J. Matls. Sci. 22, 2887 (1987).
13. A. Reisman, E. H. Nicollian, C. K. Williams, and C. J. Merz, J. Electronic Matls. 16, 45 (1987).
14. J. Blanc. Phil. Mag. B55, 685 (1987).
15. C. J. Han and C. R. Helms, J. Appl. Phys. 59, 1767 (1986).
16. G. H. Schiroky, J. Matls. Sci. 22, 3595 (1987).
17. R. R. Razouk, M. E. Thomas, and S. L. Pressacco, J. Appl. Phys. 53, 5342 (1982).
18. B. E. Deal, D. W. Hess, J. D. Plummer, and C. P. Ho, J. Electrochem. Soc. 125, 339 (1978).
19. H. Z. Massoud, Ph.D. Thesis, Stanford Univ. (1983).
20. E. A. Irene, Critical Reviews in Solid State and Materials Science, 14, 175 (1988).
21. E. A. Irene, Phil. Mag. B55, 131 (1987).
22. R. H. Doremus, to be published.
23. N. Birks and G. H. Meier, "Introduction to High Temperature Oxidation of Metals", Edward Arnold, London (1983).
24. C. K. Huang, R. J. Jaccodine and S. R. Butler in "Silicon Nitride and Silicon Dioxide Thin Insulating Films", V. J. Kapoor and K. T. Hankins, eds., Electrochem. Soc., Pennington, N.J. (1987).
25. M. Nogami and M. Tomozawa, J. Am. Ceram. Soc. 67, 151 (1984).
26. E. A. Taft, J. Electrochem. Soc. 125, 968 (1978); 127, 993 (1980); 132, 2486 (1985).
27. R. J. Charles, J. Am. Ceram. Soc. 45, 105 (1962).
28. B. Borie, C. J. Sparks and J. V. Cathcart, Acta Met. 10, 691 (1962).
29. R. H. Doremus, Thin Solid Films 122, 191 (1984).

ELECTRICAL PROPERTIES OF THERMALLY OXIDIZED SILICON

Bruce E. Deal
National Semiconductor
Fairchild Research Center
Santa Clara, California 95052, USA

ABSTRACT

Electrical properties of the thermally oxidized silicon system, including the four types of oxide charges, are reviewed. Emphasis is placed on the measurement, process dependence, and minimization and control of these oxide charges, along with possible effects on device characterization and reliability. Also discussed are oxide structure, conduction, trapping, dielectric breakdown, and surface treatment effects. Predictions of future directions and trends of thermal oxides with respect to device applications are summarized.

THERMAL SILICON DIOXIDE (SiO_2) has been used in conjunction with silicon device fabrication since the late 1950's, when scientists at Bell Telephone Laboratories used this material to selectively mask dopants during the preparation of diffused transistor structures (1). At about the same time, other Bell scientists (2-4) reported results of investigations concerning interesting properties of the $Si-SiO_2$ interface. In 1960, Hoerni of Fairchild invented the PLANAR process (5) which involved the passivation of silicon device structures with thermal oxides. This process is the basis for all silicon devices and circuits produced today. It also provided the necessary background and foundation for the invention of the integrated circuit (6,7). Since 1960, many other uses for thermal oxide layers on silicon have been developed and these are listed in Table I. One of these is as a device component, such as a gate in the MOS transistor invented in 1960 by Kahng and Atalla (8).

The kinetics of the thermal oxidation of silicon and the properties of the oxides produced have been the subject of investigation since early work at Bell Laboratories (2,3,10). The basic silicon oxidation process involves the reaction of silicon with oxygen or water vapor - typically in the 800-1300°C temperature range. The fact that the process occurs by diffusion of the oxidizing species through the oxide already formed to continuously form a new, passivated $Si-SiO_2$ interface, provides the basis for successful silicon device fabrication. Other semiconductors, such as germanium or gallium arsenide, cannot be oxidized to produce stable thermal oxides. In 1965, Deal and Grove characterized the kinetics for the thermal oxidation of silicon using a linear-parabolic model (10). An expression:

$$x_o^2 + Ax_o = B(t + \tau) \qquad (1)$$

was derived where x_o is oxide thickness at time t and A, B, and τ are constants. This

Table I
USES OF DIELECTRIC FILMS IN SEMICONDUCTOR TECHNOLOGY

- COMPONENTS IN DEVICES 1960
- CORROSION PROTECTION
- DEVICE ISOLATION
- DOPANT DIFFUSION SOURCE
- GETTER IMPURITIES
- INCREASE BREAKDOWN VOLTAGE
- INSULATE METAL LAYERS
- MASK AGAINST DOPANTS 1957
- MASK AGAINST IMPURITIES
- MASK AGAINST OXIDATION
- MECHANICAL PROTECTION
- PASSIVATE JUNCTIONS 1960
- SMOOTH OUT TOPOGRAPHY

expression is based on three consecutive reactions: (a) absorption of oxidant into outer oxide, (b) diffusion of oxidant through oxide, and (c) reaction of oxidant with silicon at $Si-SiO_2$ interface. A schematic cross-section of the thermally oxidized silicon structure is shown in Fig. 1, which also includes the designation and location of oxide charges to be discussed later. It has been shown that some of these charges are directly related to the oxidation process.

One point regarding the linear-parabolic oxidation model should be noted. While oxidation kinetics can generally be characterized by this expression over a wide range of process conditions, deviations occur in certain regions such as the initial 20 nm for dry O_2 oxidation. These deviations have been the subject of intense (and sometimes emotional) discussion and investigation since the original presentation of the model in 1965.

The remainder of this paper will be primarily devoted to a description of the properties of thermal silicon oxides, with emphasis on oxide charges and other effects important to the fabrication and operation of semiconductor devices. The latter include conductance, carrier trapping, oxide breakdown, and native/chemical oxide formation. Future trends in silicon oxidation technology will also be briefly discussed.

OXIDE CHARGES

Four general types of electrical charges have been identified in conjunction with thermally oxidized silicon structures (11-16). These charges can seriously affect device performance and reliability. Fortunately, however, their properties and dependence on processing variables are fairly well understood and their concentrations or densities, can be controlled at a minimum and acceptable level. Following is a brief description of these four types of charges, and a general picture of their origin as related to the $Si-SiO_2$ interface and the thermal oxidation process. First, however, terminology and measurement techniques are summarized.

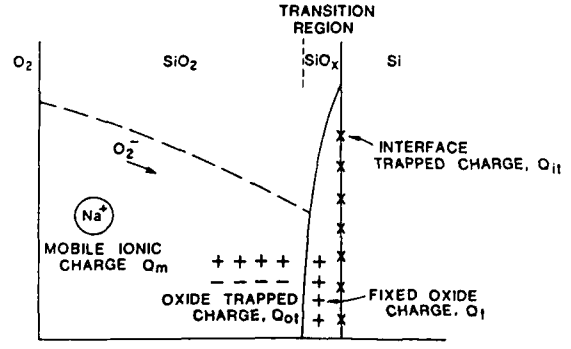


Fig. 1. Representation of mechanism and charges for silicon thermal oxidation process (10,11).

TERMINOLOGY AND MEASUREMENT - Oxide charge concentrations are reported in terms of net effective density at the $Si-SiO_2$ interface. These can be expressed as net effective charge per unit area (Q) such as C/cm^2, or as net number of charges per unit area (N) such as no/cm^2. For a given sheet of charge in the oxide, the further away from the $Si-SiO_2$ interface, the lower the net effective density. The four types of charges described below are designated as follows (16): Fixed Oxide Charge - Q_f, N_f; Mobile Ionic Charge - Q_m,N_m; Interface Trapped Charge - Q_{it}, N_{it}; and Oxide Trapped Charge - Q_{ot}, N_{ot}. A special symbol (D_{it}) is used to express interface trapped charge in terms of bandgap energy (no/cm^2-eV).

Various measurement techniques have been reported and used to determine charge densities (15). The most common and easiest to use, however, is the capacitance-voltage method (14,15). It requires only a simple MOS capacitor structure, in which aluminum dots are selectively evaporated onto the oxidized silicon and normally annealed in hydrogen at about 400°C. Typical C-V plots are shown in Fig. 2 from which charge densities can be extracted. For interface charge densities, a more complicated version of the high frequency C-V method can be used, which is called quasistatic C-V analysis (14,15). Densities of all charges are typically in the 10^9 to 10^{13} charges/cm^2 range.

Densities of most of the oxide charges can be determined by using a basic expression for the dependence of flatband voltage (V_{FB}) on various charge distributions in a thermal oxide (17):

$$V_{FB} = \phi_{ms} - \left(\frac{x}{x_o}\right)Q/Co - \frac{1}{C_o}\int_0^{x_o}\frac{x}{x_o}\rho(x)dx \qquad (2)$$

where ϕ_{ms} is the metal-silicon work function difference, Q is a sheet charge

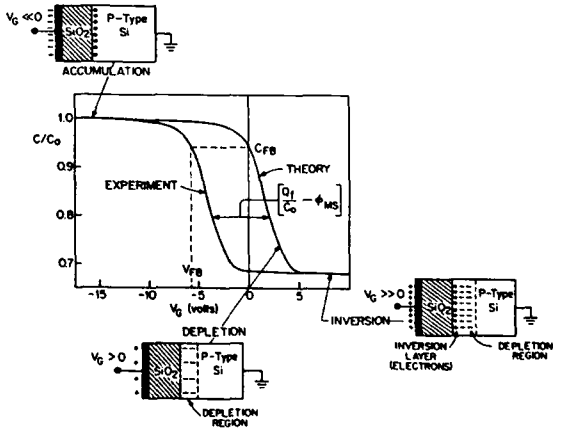


Fig. 2. Determination of fixed oxide charge density [Q_f] using capacitance - voltage technique.

distance x from the metal fieldplate, x_o is oxide thickness, Co is oxide capacitance, and $\rho(x)$ is a distributed charge in the oxide. Assuming no distributed charge, and $x=x_o$, and values assigned for various constants, the value of Q_f may be represented by:

$$Q_f/q=(-V_{FB}+\phi_{ms})(2x10^{10})/x_o(um) \qquad (3)$$

Work function differences (ϕ_{ms}) depend on the type of metal fieldplate employed and the Fermi level and dopant type of the substrate silicon. Values of ϕ_{ms} typically range from -1.0 to 1.0 V and among other things their effective values are a function of processing conditions (18).

FIXED OXIDE CHARGE (Q_f) - This charge is positive and is believed due to ionized defects (partially ionized Si) in the thermal oxide less than 20 A from the $Si-SiO_2$ interface (19). The density of Q_f depends directly on oxidation and annealing process conditions and silicon orientation [(111)>(110)>(100)]. Its density can also be changed, perhaps artificially, by electrical, chemical, and physical effects. One of its more important process dependencies is that related to final oxidation and/or annealing conditions, which is indicated in Fig. 3 by the Q_f-Q_2 triangle relationship (19). Special effects, such as impurities or nitrogen-silicon reactions can affect the "argon" curve. The origin of Q_f, as indicated later, is directly related to interface trapped charge (20) - both charges being functions of the oxidation process.

INTERFACE TRAPPED CHARGE (Q_{it}) - Unlike Q_f above, the interface trapped charge can be either positive or negative and its density will change as a function of surface potential and/or bandgap energy. Physically, it is due normally to the same defect in the oxide responsible for Q_f but is located exactly at the $Si-SiO_2$ interface. Variations of interface traps include heavy metal inclusions or radiation-induced defects. These charges can be "annealed" by active hydrogen treatment at 350-450°C, although such annealing is a reversible process. Like Q_f, Q_{it} is process and silicon orientation dependent. Its presence tends to degrade most types of device parameters. The one-to-one relationship between Q_f and Q_{it} before and after H_2 anneal is shown in Fig. 4 (20).

MOBILE IONIC CHARGE (Q_m) - The mobile ionic charge was the first to be identified and characterized (17,21). Indeed it was not possible to investigate any of the charges and instabilities in thermally oxidized silicon until the source of mobile charge was identified as impurity alkali ions (Li^+, Na^+, K^+) and their concentrations minimized to acceptable levels (below $2x10^{10}/cm^2$). These impurity ions are present in essentially every material associated with device fabrication, and many precautionary procedures have been developed to minimize their effects. Fortunately, other processes, such as gettering and masking with phosphosilicate glass and silicon nitride films, have been developed which permit the fabrication of stable, complex devices.

OXIDE TRAPPED CHARGE (Q_{ot}) - Oxide trapped charge is sometimes confused with fixed charge since their characteristics and effects are similar. Q_{ot}, however, is due

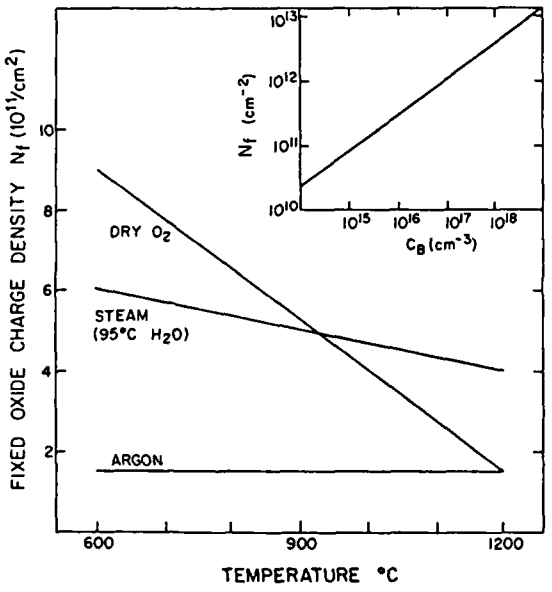


Fig. 3. Dependence of fixed oxide charge [Q_f] on final processing conditions (19).

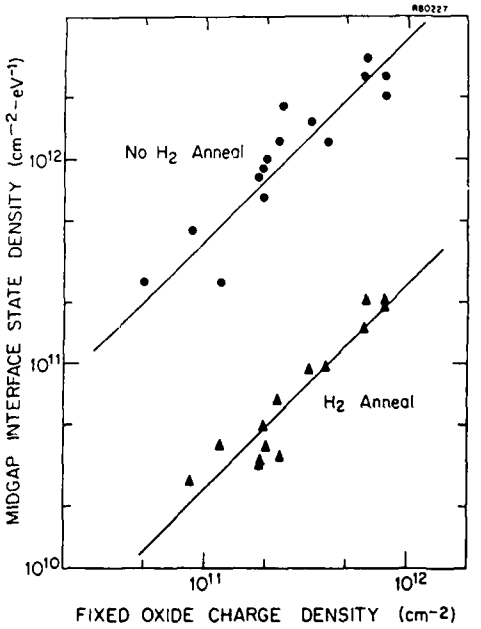


Fig. 4. Correlation between interface trap density [D_{it}] and oxide fixed charge density [N_f] (20).

to trapped holes or electrons due to either broken Si-O bonds (intrinsic) or impurity inclusions (extrinsic), and can exist in any location in the oxide (22). Often this charge is induced by ionizing radiation and the resulting trapped holes (or electrons) can be annealed at fairly low temperatures. Also, carrier trapping is brought about by high fields directed through thin oxides in certain devices. It has been reported that once a bond has been broken by a particular radiation or other type of process, the defect remains even after annealing, and will be much more suseptible to re-trapping. For this reason, some concern has been expressed regarding the many radiation-producing processes now used to fabricate VLSI and sub-micron structures. Examples are shown in Fig. 5.

ORIGIN OF OXIDE CHARGES - In Fig. 6, the proposed origin of the four types of oxide charges is presented in the form of an oxidized silicon cross-section. It can be noted that periodically an oxygen which connects a substrate silicon atom to an oxide-silicon atom is missing. This gives rise to both an oxide fixed charge and an interface trapped charge. Further into the oxide, a Si-O bond may be temporarily broken by ionizing radiation or some other electrical effect, thus permitting hole (or electron) trapping leading to the formation of an oxide trapped charge. Finally, mobile ionic charge is represented by a sodium ion, most common of the ionic impurities which diffuse rapidly through the oxide.

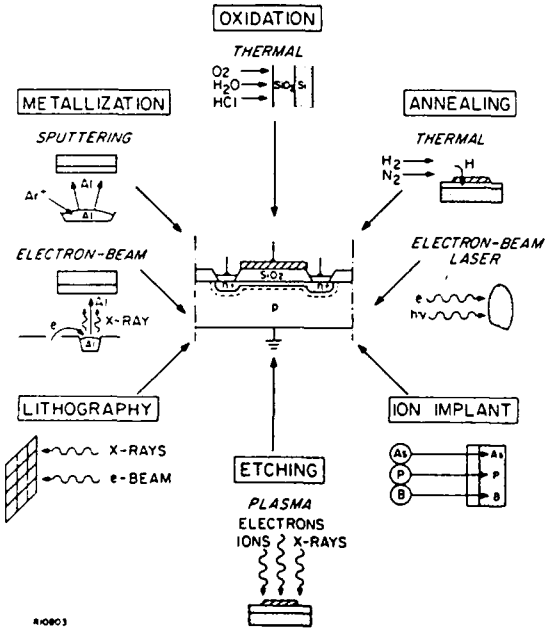


Fig. 5. Semiconductor processes which can induce oxide trapping.

THERMAL OXIDE PROPERTIES

OXIDE STRUCTURE - Crystalline quartz is structured in a regular array of six-membered silicon rings in a tetrahedral arrangement, each silicon being bonded to four oxygens and each oxygen to two silicons. Amorphous or vitreous SiO_2, on the other hand, is more random in nature, and rings may include seven or eight silicon atoms down to four near the $Si-SiO_2$ interface. The latter leads to a more dense structure which accounts for varying properties of thin oxides. Because of the more random structure of vitreous SiO_2 (formed by the thermal oxidation process) these films are more subject to defects. These defects can include distorted or broken bonds, missing oxygen or silicon atoms, or impurity incorporation (23). Such defects lead to adverse effects in device structures which include increased conduction through the passivating oxide, charge formation through hole or electron trapping, and dielectric breakdown. Some of these effects are briefly reviewed in the following discussion.

CONDUCTION - Normally thermal SiO_2 is an excellent insulator with a resistivity in the range of 10^{16} ohm-cm. It is generally believed that the basic conduction mechanism is based on Fowler-Nordheim tunneling (24). This mechanism involves electrons tunneling from the conduction band of the silicon into the conduction band of the oxide. Such a process is used, for instance, to program floating Si-gate devices by tunneling electrons thorough thin (100 A) thermal oxides at high fields. It has been found, however, that leakage or tunneling may increase for thinner oxides, which may imply that other mechanisms such as Frenkel-Poole type hopping contribute. Data obtained by Baglee and Shah (25) which demonstrate this effect are shown in Fig. 7.

TRAPPING - Carrier (electron or hole) trapping in thermal oxides has been a subject for discussion and disagreement for many years. It was first of concern to those involved in the fabrication of radiation-

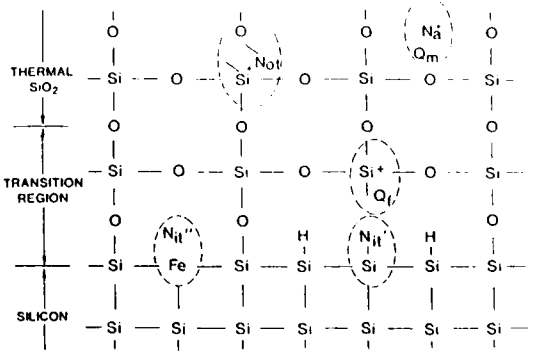


Fig. 6. Proposed origins and locations of charges in thermally oxidized silicon (16).

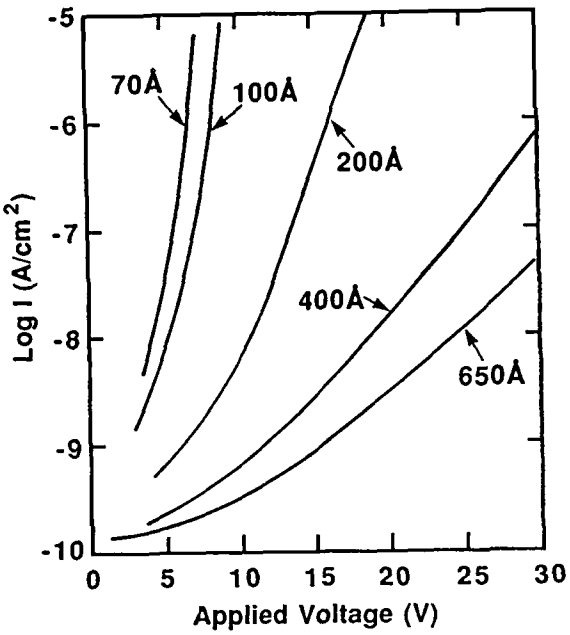


Fig. 7. Leakage in SiO_2 films of various thicknesses (25).

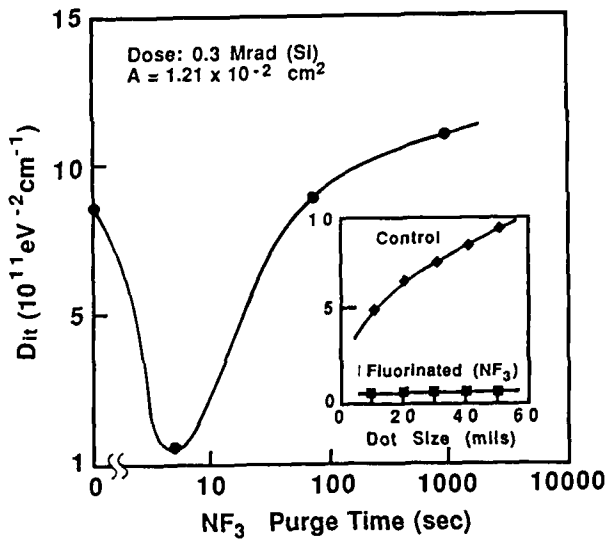


Fig. 8. Radiation-induced interface trap density as a function of NF_3 purge time during oxidation. Insert shows Al gate size effect on trap density (27).

resistant devices (26). In these investigations, it was proposed that during ionizing radiation Si-O bonds in the oxide are broken or other defects ionized (intrinsic or extrinsic) to form hole traps. Correspondingly, new interface traps are also formed at the same time. What is not agreed upon is how the hole traps are formed and how their formation is related to the associated interface trap generation. It is now fairly well accepted that such trapping is related to oxidation process conditions (temperature, ambient, impurities, etc.). One proposed approach to minimize this trapping is the addition of small, very controlled amounts of halogen species to the oxide. Data supporting this approach are presented in Fig. 8, where a minimum of electron trapping is indicated for a specific amount of fluorine added to the oxidation ambient (27).

Carrier tunneling mechanisms in various non-volatile memory devices (EPROMs, EAROMs, EEPROMs, and FLASH memories) were mentioned above. Obviously, trapping of the carriers in the device oxides is of considerable importance and concern to device engineers and manufacturers. Carrier trapping must be minimized in this case. The subject is quite complex and cannot be covered in this paper, but the reader is advised to consult various references on the subject (25,28,29).

A detailed discussion of traps in thermal SiO_2 is given by Balland (23). He indicates that traps can be characterized by physical and chemical information (nature, origin, etc.), electronic information (energy, donor or acceptor, etc.), and spatial information (density, centroid, etc.). He also lists possible origins - both intrinsic such as strained or broken bonds, silicon or oxygen interstitials or vacancies, or non-bridging oxygen, and extrinsic such as various metal, anion, alkali, or hydroxyl ion impurity additions.

DIELECTRIC BREAKDOWN - Probably of greatest concern to device manufacturers (primarily MOS devices) is thin gate oxide integrity, or stated another way, dielectric breakdown. As integrated feature sizes have scaled downwards to sub-micrometer dimensions and accordingly component densities have reached the million bit or more level, greater and greater precautions have been required during device processing to prevent shorting of oxide gates (now in the 10-20 nm thickness range) by destructive breakdown. As is the case with other failure mechanisms discussed above, the theory and mechanism of oxide breakdown is still not well understood. It is known, however, that breakdown occurs either in small, random regions in the gate due to weak spots or defective regions - probably nucleated by impurity species - or at edges of the gate region where stress is likely due to steps in the geometry. What is truly amazing is that these complex circuits can be fabricated at all with reasonable yield. One factor which may help the situation is that oxide dielectric strength apparently increases with decreasing thermal oxide thickness, as shown in Fig. 9 (25). Again, the reader is directed to the vast amount of literature on the subject of oxide breakdown (30-33).

SURFACE TREATMENT EFFECTS - Brief mention should be made of the nature and effect of

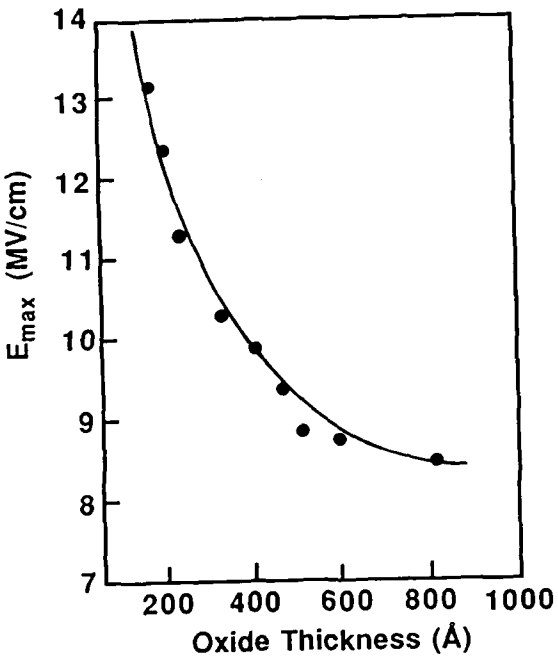


Fig. 9. Maximum breakdown strength as a function of oxide thickness (25).

silicon surface treatments on subsequent oxidation kinetics and oxide properties. Since Schwettmann (34) and Grunthaner (35) reported the effect of specific chemical cleaning procedures on oxidation kinetics, a considerable number of investigations on the subject has been carried out (36,37). The thin (0.5-2.0 nm) native or chemical oxide films produced by these cleaning steps play a critical role in other device processing areas, such as inter- connect-semiconductor contacts and poly-Si bipolar emitter structures.

CONCLUSIONS AND FUTURE TRENDS

Electrical charges associated with the thermally oxidized silicon system and other properties of thermal oxides have been reviewed with emphasis on their use in passivation and insulation in semiconductor device technology. If current trends towards submicrometer feature sizes and more complex device structures are to continue, then a better understanding of process effects and contamination control will be required. Thin SiO_2 gates may be modified as they approach 10 nm thickness with modification of composition, such as NO additions. Stress effects in oxides related to smaller and more complex device structures will have to be better modeled and controlled. If past history of semiconductor technology can be used as a guide, the above problems will be solved, and the current trends will continue.

REFERENCES

1. Frosh, C.J. and L. Derrick, J. Electrochem. Soc., 104, 547 (1957).
2. Atalla, M.M., E. Tannenbaum, and E.J. Scheibner, Bell Syst. Tech. J., 38, 749 (1959).
3. Ligenza, J.R. and W.G. Spitzer, Phys. Chem. Solids, 14, 131 (1960).
4. Atalla, M.M., A.R. Bray, and R. Lindner, Proc. I.E.E., 106B, 1130 (1960).
5. Hoerni, J.A., IRE Electron Devices Meeting, Washington, D.C., 1960; U.S. Patents 3,025,589 (1962) and 3,064,167 (1962).
6. Noyce, R.N., U.S. Patent 2,981,877 (1961).
7. Kilby, J.S., U.S. Patent 3,138,743 (1964).
8. Kahng, D. and M.M. Atalla, I.R.E. Solid State Device Conference, Pittsburgh, PA., 1960.
9. Law, J.T., J. Phys. Chem., 61, 1200 (1957).
10. Deal, B.E. and A.S. Grove, J. Appl. Phys., 36, 3770 (1965).
11. Deal, B.E., J. Electrochem. Soc., 121, 198C (1974).
12. Cheng, Y.C., Prog. Surface Science, 8, 181 (1977).
13. Feigl, F.J. in "VLSI Electronics Microstructure Science, Vol. 6," Einspruch, N.G. and G.B. Larrabee, Eds., p. 147, Academic Press, New York (1983).
14. Schwarz, S.A. and M.J. Schulz, in "VLSI Electronics Microstructure Science, Vol. 10," Einspruch, N.G. and R.S. Bauer, Eds., p. 29, Academic Press, New York (1985).
15. Nicollian, E.H. and J.R. Brews, "MOS Physics and Technology," Wiley-Interscience, New York (1982).
16. Deal, B.E., J. Electrochem. Soc., 127, 979 (1980).
17. Grove, A.S., "Physics and Technology of Semiconductor Devices," John Wiley and Sons, New York (1967).
18. Razouk, R.R. and B.E. Deal, J. Electrochem. Soc., 129, 806 (1982).

19. Deal, B.E., M. Sklar, A.S. Grove, and E.H. Snow, J. Electrochem. Soc., 114, 266 (1967).

20. Razouk, R.R. and B.E. Deal, J. Electrochem. Soc., 126, 1573 (1979).

21. Snow, E.H., A.S. Grove, B.E. Deal, and C.T. Sah, J. Appl. Phys., 36, 1664 (1965).

22. See for instance Barbottin, G. and A. Vapaille, "Instabilities in Silicon Devices - Vol. 1," North Holland, Amsterdam (1986); and Davis, J.R., "Instabilities in MOS Devices," Gordon and Breach, New York (1981).

23. Ballard, B. in "Instabilities in Silicon Devices - Vol. 1," Barbottin, G. and A. Vapaille, Eds., p. 101, North Holland, Amsterdam (1986).

24. Lenzlinger, M. and E.H. Snow, J. Appl. Phys., 40, 278 (1969).

25. Baglee, D.A. and P.L. Shah in "VLSI Electronics Microstructure Science, Vol. 7," Einspruch, N.G., Ed., p. 165, Academic Press, New York (1983).

26. For instance, see Gregory, B.L., IEEE Trans., Nuclear Sci., NS-22, 2295 (1975); Dawes, W.R. and B.L. Gregory, IEEE J. Solid-State Circuits, SC-11, 459 (1976); Reisman, A., C.J. Merz, J.R. Maldonado, and W.W. Molzen, J. Electrochem. Soc., 131, 1404 (1984).

27. daSilva, E.F., Y. Nishioka, and T.P. Ma, IEEE Trans. Nuclear Science, NS-34, 1190 (1987).

28. Verma, G. and N. Mielke, Proc. IEEE Rel. Phys., p. 158, Monterey, CA. (1988).

29. Garrigues, M. and B. Balland in "Instabilities in Silicon Devices - Vol. 1," Barbottin, G. and A. Vapaille, Eds., p. 441, North Holland, Amsterdam (1986).

30. Harari, E., J. Appl. Phys., 49, 2478 (1978).

31. Hokari, Y., T. Baba, and N. Kawamura, IEEE Trans. Electron Devices, ed-32, 2485 (1985.)

32. McPherson, J.W. and D.A. Baglee, J. Electrochem. Soc., 132, 1903 (1985).

33. Holland, S., I.C. Chen, J. Lee, K.K. Young, and C. Hu, in "Silicon Nitride and Silicon Dioxide Thin Insulating Films," Kapoor, V.J. and K.T. Hankins, Eds., p. 361, The Electrochemical Society, Inc. (1987).

34. Schwettmann, F.N., K.L. Chiang, and W.A. Brown, Abs. 276, p. 688, The Electrochemical Society, Seattle, WA (1978).

35. Grunthaner, F.J. and J. Maserjian, IEEE Trans. Nucl. Sci., NS-24, 2108 (1977).

36. Becker, D.S., W.R. Schmidt, C.A. Peterson, and D.C. Burkman, in "Microelectronics Processing: Inorganic Materials Characterization," Casper, L.A., Ed., p. 366, American Chemical Society, Washington, D.C., 1986.

37. Ruzyllo, J., G.T. Duranko, and A.M. Hoff, J. Electrochem. Soc., 134, 2052 (1987).

INTERLEVEL DIELECTRICS AND PASSIVATING FILMS

Geraldine Cogin Schwartz
IBM East Fishkill
Hopewell Junction, New York, USA

TO MAKE THE FULLEST USE of the increased speed of the smaller devices, e.g., transistors, that can now be built within the semiconductor and to make the fastest and most efficient connections among them, multiple wiring planes are used (1). Multilevel metallization reduces the length of the interconnections and the capacitive coupling to silicon, thereby decreasing the wiring delay. Interlevel dielectric layers are needed to isolate the several metallic conducting planes from each other; these films also insulate the conductors within a plane. An example of a mutlilevel bipolar structure is given in Fig. 1 (2).

The wiring delay is directly proportional to the resistivity (p) of the interconnect, the square of its length (l), and the dielectric constant (e) of the insulator; it is inversely proportional to the thickness (t) of the insulator. Therefore, if possible, the insulator of choice would be a very thick film of a material whose dielectric constant is low.

Polyimides (the only class of organic films, now available, whose thermal stability is compatible with interconnection processing temperatures) have a dielectric constant that is lower than that of inorganic insulators. However, as will be discussed below, there are several disadvantages to their use; nevertheless they have been studied extensively and their use is growing.

There is a limit to the thickness of the interlevel dielectric layer. As the insulator thickness is increased, the capacitance between conducting planes is decreased. But, since the thickness of the metal covering the insulator must also increase, the capacitance between conductors on a given level in increased. Another problem, which arises when the dielectric/metal thickness is increased, is one of processing. As the metal thickness is increased, so must the next level of insulator. It becomes increasingly difficult to cover the metal and fill the spaces between them, without voids, as the aspect ratio, AR, (step height/space) increases, as shown in Fig. 2a (3). Solutions to the problem of filling these spaces (gaps) will be addressed in the section on planarization.

Another processing problem, which may also limit the insulator thickness, is etching holes in the insulator.

Contact holes are needed to connect elements within the semiconductor (e.g., the base, emitter, etc. of a bipolar transistor) or elements of a field effect transistor (FET) (e.g., a gate) to the metal interconnects in the next plane. Via holes interconnect the several metallized planes. But there may be a practical limit to the depth of the hole that can be etched. Advances in lithographic systems have made possible alignment and printing of small features. Accurate pattern transfer can no longer be accomplished by wet chemical etching. Instead, reactive plasma assisted etching (called reactive ion etching, RIE, reactive sputter etching, RSE, etc) is now used almost universally (4). However, in these processes, the mask is eroded as the hole is etched. Since the acuity of printing the images degrades as the resist thickness is increased, there is a limit to the thickness of the insulator which can be etched, unless more complicated masking schemes, such as multilevel resist masking (MLR), is used (5).

Another limitation arises because the etched holes must be filled with metal. In order to pack a very large number of devices into a chip, the contact and via holes are made as small and as steep as possible. If the insulator is very thick, these holes have a very high AR; this is the problem illustrated in Fig. 2a. It should be noted that the difficulty in covering steps and filling holes also increases as the wall angle increases; this is illustrated in Fig.2b (3).

The materials chosen as interlevel dielectrics must satisfy other requirements in addition to a low dielectric constant. Some of these are listed below, together with some desirable values: High break-down strength (10E7 V/cm), low conductivity (10E17 ohm-cm), low compressive stress (<1E9 dynes/cm2) (crack resistance), low defect density (<0.1/cm2), low levels of impurities, good adhesion, stability toward the environment and processing chemicals, resistance to penetration by moisture (high density) and by mobile ions, and etchability.

Film deposition processes must be reproducible

and compatible with the underlying materials and structures. The films should be uniform in thickness (+/-2%). Coverage of underlying topography should be conformal or planarizing. For semiconductor manufacturing, they must also be economically acceptable; throughput should be high whereas capital expenditure and the clean room floor space required should be low.

No single material or deposition process satisfies all the requirements; compromises must be made.

Polyimides have already been mentioned as an interlevel insulator. The others in common use are inorganic. Silicon dioxide (SiO2) and silicon nitride (Si3N4 and plasma nitride) are used most frequently, silicon oxynitrides to a limited extent.

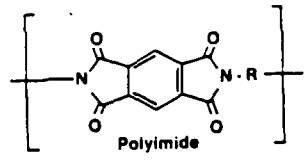
Many techniques must be used to characterize dielectric films. Too often, a single one is used and wrong conclusions may be reached. Some of the techniques used determination of chemical, structural, and mechanical properties are given below.

The chemical composition of the inorganic insulators has been determined by Rutherford (nuclear) backscattering (RBS), Auger electron spectroscopy (AES), secondary ion mass spectroscopy (SIMS), and microprobe analysis. The hydrogen (H) content has been measured by proton-proton scattering (6a,b), elastic recoil detection (EDR) (6c), and by resonant nuclear reaction techniques (7), which, for silicon nitride, have been correlated with the infrared (IR) absorption bands of Si-H and N-H (8). Among the more commonly used optical techniques for the determination of refractive index/dispersion and film thickness are VAMFO, variable angle monochromatic fringe observation (9), CARIS, constant angle reflection interference spectroscopy (10), the prism coupler (11) and, in certain thickness ranges, ellipsometry. This last technique is the most sensitive to any defects, such as roughness, in the underlying substrate. The refractive index of a film is influenced by its chemical composition, density and impurity content; its value cannot be used, alone, for characterization. For inorganic films, the etch rate in buffered HF or preferential etches such as "P" etch (12) gives an indication of composition, bond strain and density. IR spectroscopy provides compositional information; it has been particularly valuable in evaluating the stoichiometry, density, porosity, as well as the H2O, OH, and impurity content of silicon oxides (13) and, as mentioned above, the SiH and NH content of nitrides. Moisture absorption by polyimides has been determined by weight increase and change in dissipation factor, but these methods are relatively insensitive; more sensitive are surface I-V characteristics, mass spectrometric data, and capacitance measurements. Moisture permeation studies of polyimide films have also been made (14). Density can be measured by weighing or by the use of a density column. The porosity of inorganic insulators has also been determined by the densification which occurs upon heating (thickness/refractive index changes) (13a). Stress is determined by measurement of wafer bowing (using a fiber optic probe, a profilometer, or the X-ray lattice curvature) or by the use of a cantilevered beam.

The important electrical properties, for interlevel insulators, determined by standard methods are: resistivity, dielectric constant/dissipation factor, and break-down strength.

POLYIMIDE

Polyimides are a class of organic polymers whose generic formula can be written:



Polyimide

Polyimides, with a variety of electrical and mechanical properties, are available commercially (15). The stress in polyimide films usually is low. Polyimides vary in flexibility and in their ability to flow and fill small gaps. The coeffient of thermal expansion is about about one to two orders of magnitude greater than that of silicon and the inorganic insulators, although a new low-expansion polyimide has been reported (16).

Polyimides have been used alone or in a composite layer with inorganic films as interlevel dielectrics in multi-level semiconductor devices. Wilson (17) and Samuelson (15b) have reviewed the use of polyimide for VLSI.

One of the most desirable properties of polyimides is a low dielectric constant, e=2.9-3.7 compared to SiO2, e=3.8-4.2, which has a lower dielectric constant than other inorganic insulators in common use (e.g., silicon nitrides, Al2O3, etc). However, upon exposure to air, polyimides readily absorb moisture; this increases the dielectric constant, dissipation factor, surface leakage, and polarizability (14). Thus, the interlevel insulator in the finished product may not have the low dielectric constant and other desirable electrical properties of the pure material. Water absorption may depend on the structure of the polyimide (14).

Another attraction of polyimides is the low cost of deposition. It is applied by spin-coating, which makes unnecessary the use of vacuum or other expensive and space-consuming apparatus required for deposition of inorganic insulators. A solution of the pre-polymer is applied to a wafer. A series of heating cycles expels the solvent and ultimately converts the pre-cursors to polyimide; this is called the cure cycle. The chemical reactions, which occur during curing, have been described in the literature (18). Application is rapid, but the cure cycles are time-consuming and require careful control since both the amount of solvent remaining after bake and the extent of chemical reaction determine the electrical and mechanical properties, the solubility, swelling, lifting, and crazing upon exposure to solvents, as well as the reliability of the polyimide. The shelf-life of the pre-cursor solutions is an important factor since viscosity changes, affecting film thickness, can occur during storage. Vacuum deposition of a polyimide film having a very low dielectric constant (2.9) has been reported (19);

this low value may make the material attractive despite the higher cost of application.

Initially, interest in polyimides was ignited by reports that the material planarized underlying topography. However, several investigators have shown that the degree of planarization depends on the type of polyimide, the solids content, the molecular weight, and on the dimensions of the lines and spaces of the underlying pattern. Thin lines separated by narrow spaces were planarized more competely than wide lines; groups of closely spaced lines acted a one wide line (15a). Planarity was improved by using several coats of polyimide, curing between layers (15a) and by modifying the curing method (20). However, complete planarity has not been achieved for all features. These observations have been verified in the study of planarization of other kinds of spin-on organic films as well (21). Although planarization is limited, polyimides do alter the edge profiles of the underlying patterns; steep steps are made more gentle so that the edge coverage of the overlying metal is improved; the slope angle decreases as the thickness of the polymer film is increased (20b,21b).

A disadvantage to the use of polyimide is its poor adhesion to aluminum (Al) and Al alloys, the most widely used conductors, and to some inorganic insulators as well as to itself (22). An adhesion promotor or plasma ashing is often required. Adhesion of Al to polyimide is also poor. Adhesion layers, such as thin films of chromium, have been interposed between the Al conductor and the polyimide. The poor adhesion to Al, together with the hygroscopic nature of polyimide, results in increased sensitivity to metallic corrosion. Nishida (23) reported that corrosion was caused by moisture within the bulk of PIQ polyimide. The moisture content, coupled with poor adhesion, allows the formation of a monolayer film of water said to be necessary for corrosion initiation (24).

The release of moisture, trapped within the polyimide film or at the film/metal interface, was reported to be responsible for the formation of bubbles within a polyimide film (25). Once a metal pattern is formed on the surface of a polyimide film, the only path for dehydration is at the edges of the metal; dehydration is, therefore, a lengthy process.

The interlevel (via) resistance of structures formed by plasma etching via holes in polyimide was reported to be exceptionally high (26). This was attributed to re-deposition or plasma polymerization of etch products. In-situ sputter cleaning has been used to lower via resistance by etching the insulating surface metallic oxide and preventing its regrowth. However, sputter cleaning a polyimide surface creates a conducting layer, which must be removed before further processing.

According to Wilson (17), the pin-hole density of polyimide films is equivalent to or lower than that of inorganic insulators. However, the electrical properties (e.g., resistivity and breakdown voltage) of polyimide films are inferior to those of SiO2 films; multiple films of polyimides have better electrical characteristics than a single layer of the same thickness (27). The electrical conductivity (bulk electronic, ionic, and surface/interface) of polyimides has been studied by Brown (28).

The thermal conductivity of polyimide is lower than that of inorganic insulators; heat transport may become a limiting factor in the use of this material (29).

There have been reports (29) that polyimide passivation of AlCu conductors resulted in improved electromigration resistance, but another study (30) found no difference.

Some of the disadvantages of polyimide have been overcome by the use of a composite film of polyimide and an inorganic material. However, depending on the final composition, there may be a significant increase in dielectric constant; in addition, the processing may no longer be simple or cost-effective. For example, Eggers (22a) suggested coating polyimide with a thin layer of plasma nitride (to be described below), being careful to dehydrate the polyimide by in-situ heating. The nitride cap served as an etch mask in forming the via holes, improved adhesion, and further reduced the defect density, but the processing was more complex. Examples of a dual polyimide/oxide structure have also been described (16,31)

The processing steps used for polyimide interlevel insulators are often similar to those used for inorganic, usually SiO2, insulators (31). The resulting structures are similar, except for some degree of planarization and the improvement of step coverage mentioned earlier. Planarizing procedures will be described in a later section. The introduction of photosensitive polyimides (32) have the potential for process simplification, but thus far, the speed and resolution of the materials are inadequate. Etching of small holes in polyimide is accomplished by RIE, usually in oxygen. The shape of the holes depends on the mask and on the etch parameters.

INORGANIC INSULATORS

The most widely used processes for depositing inorganic insulating films are: (i) thermally activated chemical vapor deposition, CVD, performed at atmospheric pressure (APCVD) and at low pressure (LPCVD), (ii) plasma-enhanced CVD (PECVD), and (iii) rf sputter deposition. These will be discussed below. Other techniques, such as electron-beam (33a), laser (33b), and nozzle-beam (33c) evaporation, ion-beam deposition (33d,e) and photo-enhanced CVD are not used extensively. However, advances in laser technology (33f,g), and vacuum ultraviolet light sources (33h,i) have re-kindled interest in photochemical deposition processes. Spin-on glasses, SOG, (33j) e.g., polysiloxanes, silicas, silicates in organic solvents) are potentially attractive because of the low cost of application. However, they have been used used only as sacrificial layers (as discussed later) or in thin layers in combination with other more stable insulators, because thick layers crack during cure, i.e., during conversion to SiO2 (33k).

The inorganic materials used as interlevel dielectrics are amorphous films.

SiO2 is used, with and without doping by phosphorous (P) and/or boron (B) or germanium (Ge). Also used, as mentioned previously, are silicon

nitride (Si3N4 and plasma nitride) and silicon oxynitrides. Aluminum oxide (Al2O3) had been used in the past because it is possible to fabricate planar structures by selective anodization of aluminum alloys (34); the high dielectric constant of Al2O3, process complexity, and the development of new techniques of planarization, metal lift-off (2,35), and plasma-assisted etching (36) led to the abandonment of this process.

CHEMICAL VAPOR DEPOSITION (CVD)

Film formation by CVD is a heterogeneous reaction: volatile reactants produce a solid film upon reaction at a surface. The sequential kinetic steps in the heterogeneous CVD process have been described: (i) transport of reactants to the substrate, (ii) adsorption of the reactants on the substrate surface, (iii) chemical reaction (including surface diffusion) and nucleation on the surface, (iv) desorption of the volatile products from the surface, and (v) transport of the gases away from the surface (37). Many CVD reactors are available commercially and have been described in the literature (37a,38). They can be classified as hot-wall (substrate and reactor walls at same temperature; deposition on all surfaces) or cold-wall (substrate at higher temperature than walls; deposition only on sustrate) systems.

As mentioned above, CVD depositions are carried out either at atmospheric pressure (APCVD) or at about 1 Torr (LPCVD). At higher pressure, the mass tranfer rates of the volatile reactants and products are of the same order of magnitude as the surface reaction rate. Reducing the pressure greatly enhances the mass transfer rate; the surface reaction becomes the rate-limiting step. Mass transport depends on reactant concentration, boundary layer thickness, and diffusivity; these are related to the reactor configuration, flow rates, etc. In LPCVD, reactor design is less critical since the rate limiting step (i.e., the surface reaction rate) depends chiefly on reactant concentration and temperature (38a). Uniform deposition is achieved more easily in LPCVD but at a sacrifice of reaction rate.

The chemical composition, electrical and mechanical properties, as well as the uniformity and deposition rate of the films are controlled by many factors. Among them are: substrate and reactor-wall temperatures, reactant ratios and flow rates, use of diluents and their flow rates, reactor configuration, surface and reactor cleanliness, surface topography, and possibly other factors. One must consult the literature for a detailed discussion of the interactions (37,38) since they are complex and difficult to catagorize and describe succinctly.

Homogeneous gas phase reactions are responsible for formation of dust which results in hazy films. These films often have a high defect density and poor dielectric properties. Therefore, such reactions must be suppressed.

Good quality CVD films are produced only at relatively high temperatures (>500°C). Lower temperature CVD processes have been developed, but the film quality usually is inferior or has not been reported. Therefore, CVD films are most commonly used only when the underlying conductors are refractory materials, such as silicon, tungsten, or the refractory metal silicides. Hence, CVD films are usually used in FET devices, over refractory gates and/or interconnects; in bipolar devices, the higher resistivity of such interconnects is unacceptable.

SiO2 films are used very widely; undoped SiO2 is preferred, where possible, because of its lower dielectric constant.

CVD SiO2 films are usually formed by oxidation of silane (SiH4) or chlorosilanes by O2 or nitrous oxide (N2O) (39a) and by the decomposition (pyrolysis) (39b,c) or oxidation of tetraethylorthosilicate (TEOS). Ozone has been used as the oxidant of TEOS (39d). The advantage to the use of TEOS is its higher surface mobility (40); it can migrate over the entire surface, into irregularities, before the heterogeneous reaction occurs on the surface. Phosphine, trimethylphosphite, diborane, boron trichloride, trimethyl borate, triethylborate, or germane are among the compounds added to the reaction mixture for doping by P and/or B (41a-c) or Ge (41d). Doping SiO2 with P (to form PSG) makes it a barrier to alkali ion penetration; this is important in FET devices which are sensitive to such contamination. Addition of P and/or B or Ge to SiO2 lowers the temperature at which the oxide can be re-flowed for tapering or planarizing underlying topography. The flow temperature is too high for this technique to be used with Al-based conductors; it is restricted to FET devices in which refractory metals are used. CVD oxides, particularly those deposited at lower temperatures, are tensile; the stress can be reduced and can even become compressive at high concentrations of P. However, films doped very heavily with P are hygroscopic; this can cause current leakage across the film surface and metallic corrosion.

CVD Si3N4 is a barrier to alkali ion migration and it is used primarily for this purpose. It is deposited before the conducting metallurgy. Si3N4 has less of a thermal mis-match with silicon than does SiO2.

CVD Si3N4 is formed by the reaction of SiH4 or chlorosilanes wih NH3 or NH3/N2 mixtures. The H-content of the films depends chiefly on the deposition temperature (42). The high dielectric constant of Si3N4 makes it unattractive for use in high performance devices; the high deposition temperature makes it unsuitable as an interlevel dielectric between Al or Al alloy interconnection levels. The films are tensile and crack when their thickness exceeds about 3000Å.

Step coverage of deposited films is almost never conformal because of shadowing of a multi-directional source. The step coverage of films deposited by APCVD is different from that of films deposited by LPCVD, as shown in Fig. 3 (40). Enhanced surface migration improves step coverage; since migration increases with increasing temperature, so does step coverage. The higher surface mobility of TEOS has already been mentioned.

PLASMA-ENHANCED CHEMICAL VAPOR DEPOSITION (PECVD)

In PECVD, a glow discharge provides the energy for the chemical reaction. In these non-equilibrium plasmas, the electron temperature is high, the gas temperature low. Electron-impact-dissociation in the discharge results in the formation of reactive species (largely free radicals) that normally are formed at high temperature. This allows the formation of films of relatively high quality at low (~400 °C) substrate temperatures. Because of the relatively low deposition temperature, PECVD insulating films are used extensively as interlevel dielectrics. Hess (43) has pointed out a further advantage of PECVD: the highly reactive plasma atmosphere can result in the formation of materials with unique chemical, physical, and electrical properties.

PECVD films usually contain large amounts of hydrogen, although in some deposition systems, films with low or negligible H-content have been produced. The effects of hydrogen will be discussed below. The growing film is subjected to bombardment by energetic species which can influence film properties (43,44). However, another consequence can be radiation damage, some of which can be annealed out only at temperatures too high to be compatible with other materials or structures in the device. In some reactors this exposure has been minimimized or eliminated.

PECVD systems are more complex than CVD systems because of the need for rf generators and their associated networks. In conventional PECVD systems, the range of frequencies is 50 kHz to 13.56 Mhz and the range of pressure is about 50 mTorr to 5 Torr. Microwave excitation has also been used, and in a newer configuration, discussed below, much lower deposition pressures have been feasible.

There are several types of PECVD reactors. The first work was carried out in bell jars or in a tubular (barrel) reactor powered by external rf coils or plates (45). Although fundamental studies were possible, deposition was not uniform, and they were unsuitable for commercial application. The earliest practical commercial system was the radial-flow, capacitively-coupled parallel plate diode reactor with internal electrodes (46). Uniformity was improved significantly, and the use of PECVD films in the semiconductor industry grew rapidly. There are many variations of this kind of system available commercially (47). A PECVD diode reactor using magnetic enhancement of the glow has been described (48). New designs in hot wall furnace (tubular) batch reactors with internal electrodes have been introduced as competitors to the (cold wall) parallel plate systems (49). All of these reactors are batch systems with moderate (250-700 Å/min) deposition rates. Defect densities are relatively high. System cleaning is a problem; deposits are formed on all surfaces and can flake when they get too thick or are exposed to the atmosphere. The use of load-locks diminishes the latter problem. Some of the newest reactors, such as the single wafer, multiple-chamber (50), and a multiple-station reactor (51), use in-situ plasma cleaning as well as load-locks. They are high rate systems; deposition rates of 5400 Å/min in the multiple-station reactor have been reported (51). Therefore, their through-put can be competitive with batch systems. In these systems, the time between loading and deposition is short so that hillock formation on Al interconnects is reduced.

In the reactors discussed above, all the reactants, such as SiH4, TEOS NH3, N2O etc. (depending on the film to be deposited) are activated in the glow; the resulting films often contain significant amounts of hydrogen. In oxides, the concentration of H can be 5-10%rbl.at.%; some plasma nitride films contain as much as 30 at.% H. However, it has been reported that dilution of an SiH4/N2O mixture with helium (He), reduced the H-content of the resulting oxide film and also reduced the oxygen deficiency which resulted from reactant ratio used (52). These results, obtained using a standard parallel plate reactor, are similar to those obtained in the remote plasma CVD (RPECVD) reactor discussed below.

In RPECVD (53), only the N2O, N2, or NH3 (usually diluted with He) are activated by an rf plasma and the excited species transported out of the glow to react with (unexcited) silane to form a gaseous precursor. When this precursor undergoes a heterogeneous (CVD) reaction on the surface to form the desired film, H is eliminated. For certain deposition conditions, the H content of RPECVD films can be very low; a H-content below the detectibility limit of IR spectroscopy, i.e., 1%, has been reported (53f). Another advantage of this technique is that the substrates are not exposed to the glow discharge. The original system was a single wafer reactor, and the deposition rates were very low (~100 /min). A high rate (1500-2000 /min) adaptation of this idea, which has become available commercially, uses a microwave discharge for excitation (54).

Another version of RPECVDC, ECR deposition, employs a microwave electron cyclotron resonance plasma for very efficient excitation of the O- or N-source gases (55). The plasma is extracted from the discharge chamber using a divergent magnetic field. The growing film is bombarded by ions whose energy is about 20 eV. ECR deposition is carried out at a pressure of about 1 mTorr. Impingement of the active species is normal to the wafer surface; therefore, filling high-aspect-ratio spaces is more feasible. In other methods, the angle of incidence is oblique; this makes gap filling more difficult. Deposition rates of about 200-500 /min are reported (55,56); the use of an rf bias at the substrate (to be discussed in the section on planarization) reduces the rate significantly. Some claim that films deposited at low substrate temperature using ECR are superior to those deposited by other PECVD techniques (55a,b). Plasma nitride films containing relatively low concentrations of hydrogen (<5%) have been prepared using ECR (55c).

A gradient field PECVD (57) reactor was designed to deposit plasma nitride films with a low level of radiation-induced damage.

Induction-heated (IH) PECVD of SiN films has been reported (58); it has been used to deposit low H-content plasma nitride in two modes: one is a photo- and radical-assisted mode used to eliminate

radiation damage; the other, a higher rate PECVD mode.

Among the films used most extensively are the so-called plasma nitrides; they are often referred to as SiN although SixNyHz is often a more precise description. They are formed by the reaction of SiH4 with NH3 and/or N2. The stoichiometry of the films depends on the deposition conditions; the ratio SiH4/NH3 or SiH4/N2 is the primary determinant. The films are most commonly Si rich, i.e., Si/N> 0.75, although N-rich films have been prepared (55c,d,59). Near-stoichiometric Si3N4 films have been deposited, under certain conditions, using REPCVD (53h) and ECR (55c,d). The H-content of PECVD films has been mentioned previously. Both Si-H and N-H bonds can be formed in SiN; the ratio between them depends on deposition conditions. The H-content influences the density, stress, and etchability in HF-based etchants. It also plays a role in determining the electrical properties of the film. Out-diffusion of hydrogen has been reported to cause instabilities in SiN-passivated FET devices (60a,b). Upon heating to temperatures of 400-650 °C, plasma nitride films undergo a rapid stress relaxation which has been attributed to the evolution of unbonded hydrogen (60c). The lower the H-content, the better the film. Impermeability to Na+ is an important property on SiN films.

Plasma silicon dioxide has been deposited most frequently by the oxidation of SiH4 by N2O (45c,61), O2 (62), and CO2 (47c), and by the decomposition or oxidation of TEOS (45a,b,63). Plasma silicon oxide is nearly stoichiometric SiO2. Both Si-H and OH bonds have been detected in most films. Sometimes P- and/or B- containing dopants are added, for the reasons given above. Because the dielectric constant of SiO2 is about half that of SiN, it is most widely used as an interlevel dielectric.

Silicon oxynitrides, in which the electrical properties of SiO2 are combined with the physical properties of silicon nitride, are used to some extent. The ratio of the nitrogen to the oxygen in the film will depend on the ratios of the source gases and will influence the properties of the film (64). Conventional parallel plate (64a), hot wall (64b), and ECR (64c) reactors have been used to deposit these films.

The properties of the deposited films depend on a wide variety of process parameters: power, pressure, flow rates, reactant concentration and reactant ratios, diluents, substrate temperature, electrode spacing, and rf frequency as well as on the reactor configuration which influences film uniformity to a great extent. Consult the literature for a more detailed and specific discussion of these interactions (47,65).

Step coverage of PECVD films has been modelled by Ross and Vossen (66); their results are shown in Fig. 4. It is clear that the step coverage is not conformal in this analysis although some improvement over LPCVD might be expected because of the possibility of redistribution by ion bombardment. Step coverage improves as the wall angle and aspect ratio decrease. Higher surface temperature, which increases surface mobility, also improves step coverage. The step coverage of PECVD TEOS films has been studied experimentally (63c).

RF SPUTTER DEPOSITION

Sputtering is a physical process (67). When ions whose kinetic energy exceeds the binding energy of the surface atoms of a solid strike the solid, atoms are ejected. Sputtering is a result of a collision cascade, a sequence of independent binary collisions; it is not a simple interaction of an incoming ion with a surface atom. The solid from which the atoms are ejected is called the target. The composition of the film deposited on the substrate is usually the same a that of the target, although oxygen and nitrogen deficiencies have been reported. The sputtered material is usually monatomic although polyatomic species, e.g., SiO have been detected (68). Most of the energy of the ions incident upon the target is transferred to its surface as heat. The temperature of the substrate is raised by bombardment by secondary electrons produced at the target.

Insulators are sputtered by positive ions formed in an rf glow discharge; argon (Ar) is used most commonly, at a pressure of about 1 to 100 mTorr. If a dc potential is applied directly to an insulating surface, a positive charge accumulates, preventing further positive ion bombardment. To avoid this, the most useful technique is to apply an rf potential to a metal electrode behind the insulating target. A frequency of >100 kHz is required to sustain a continuous discharge. Most sputtering systems operate at 13.56 MHz. Higher frequencies (multiples of 13.56MHz) have been used because, as the frequency increases, the ion current density increases but the ion energy decreases (69). Therefore it is possible to achieve higher deposition rates at lower target voltages. Reducing the target voltages reduces the energy of the secondary electrons produced at the target. Damage to FET devices is probably caused by production of X-rays generated in the gate by these secondary electrons (70); reducing their energy reduces the damage (71a). In addition, substrate heating is decreased. Another method of reducing damage has been the introduction of a dc-biased mesh grid between the target and substrate (71b).

Insulators can also be deposited by reactive sputtering of a metallic target in a suitable reactive gas mixture, such as O2/Ar, N2/Ar, N2, in a dc or an rf discharge (72).

Among the advantages of sputtering are: (i) controlled stoichiometry of the deposited films, (ii) easy sputter cleaning of substrates, (iii) improved adhesion and film density, (iv) bias sputtering for manipulation of film properties, planarization and gap-filling, (v) thickness control, and (vi) low temperature deposition possible. Among the disadvantages are: (i) low deposition rates, although high rate systems are now being developed (73), (ii) the need for cooling the target and, possibly, the substrate, (iii) system complexity since both high vacuum apparatus (the presence of residual gases can affect deposition rates and possibly film properties) and rf networks are required, and (iv) contamination from the sputtering chamber and by gas incorporation. The former arises from two sources: (a) sputtering of the walls (see below) and (b) flaking of the film deposited on the

interior surfaces, particularly when the system is opened to the atmosphere; both can affect film quality and reliability. The latter does not always have deleterious effects; e.g., large amounts Ar can be incorporated into sputtered SiO2 films (Ar/SiO2=/<0.1) without any adverse adverse effect on their insulating properties (74).

A typical rf sputtering system consists of (i) a vacuum chamber and its associated vacuum systems, (ii) the electrodes, (iii) the rf power supply, and (iv) the matching network between the glow discharge and the generator.

In a two-electrode (diode) system, the target electrode (with an insulator on its surface) is the smaller electrode; the substrate electrode is connected to the grounded chamber. This configuration is used because the smaller electrode is the one at which the higher negative dc bias is developed (75). The plasma develops an appreciable positive potential with respect to both electrodes; therefore the substrate and chamber are also negatively biased and bombarded by positive ions. The ratio of the voltages at the electrodes is inversely proportional to the electrode areas. To minimize sputtering of the chamber walls, the area ratio, (substrate + chamber)/target, is made as large as possible.

It has been shown conclusively that the dielectric and physical properties of insulators (such as dielectric constant/dissipation factor, resistivity, breakdown voltage, etch rate in HF-based etchants, pinhole density, coverage of underlying metal stripes) improves with increasing substrate bias (76). The improvement with increasing bias has been shown to be due to an increase in the re-emission coefficient, R, the fraction of impinging particles re-emitted from the collecting surface (77). R is related to the binding energies of species deposited on the surface. The improvement in quality with increased substrate bias is related to the removal, by sputtering, of atoms trapped in non-optimal surface sites. Raising the surface temperature also increases R.

Therefore, three-electrode (triode) systems, in which the substrate bias can be controlled (without substantial sputtering of the chamber walls), have been introduced. In these systems the chamber is grounded and the target and substrate electrodes are isolated. One example is the tuned system (78a), in which the bias on the substrate is controlled by the use of a tuning network, another is the driven system (78b), in which both electrodes are powered by the use of two rf generators or a power-splitting network and a single power supply.

Magnetrons are diode sputtering systems in which magnetic fields are used to produce high ion currents (for high deposition rates) at low voltage (low electron energy), but configured so that the discharge is uniform. They are operated at lower pressures than the other kinds of sputtering systems. Many configuration have been described (68d,79).

SiO2 is the most widely used sputtered insulator. Its low dielectric constant (compared to the other inorganic dielectric films) has been mentioned. It is sputtered from a dense target of very high purity silica. The deposition rate (or, more accurately, the rate of accumulation of the film on the substrate, since some of the deposited material may be re-sputtered or re-emitted thermally; however, deposition rate is used used almost universally) depends on several parameters. It increases with increasing current/peak-to-peak voltage. At low pressures the rate increases, due to an increase in sputtering species; at intermediate pressures the rate is insensitive to pressure; at higher pressures, the rate decreases because of back-scattering. The rate decreases with increasing target/substrate distance (back-scattering) and with increasing temperature. Uniformity is a function of the target/substrate separation.

Sputtered SiO2, deposited under optimal conditions, i.e., with adequate re-sputtering, has high resistivity and break-down strength; it is dense and a good moisture barrier. The films are very nearly stoichiometric in composition; they contain minimal amounts of H and OH. Although deposited at lower temperatures than CVD films, the stress of sputtered SiO2 is compressive. The stress in the film increases with increased re-sputtering. Because of reduced thermal mismatch, the room temperature stress is lower when the films are deposited at lower temperatures, but this is at the expense of other properties, which, as in most deposition techniques, improve with increasing temperature. These films, as mentioned earlier, contain significant amounts of Ar; those with high Ar content resemble thermal oxide most closely. The compressive stress in the film is thought to be a consequence of Ar incorporation. Small partial pressures of O2 in the Ar decrease the deposition rate sharply; the addition of H2 or He to the Ar causes a very small decrease.

Sputtered Si3N4 has some properties that are superior to CVD Si3N4. The deposition temperature is lower, and its stress is compressive. The H-content of sputtered films is negligible. However, direct sputtering of a Si3N4 target is not used extensively. The targets are hot pressed or sintered; although they are of high purity, they contain trapped gase which are released and incorporated into the film. When Ar is used as the sputtering gas, dissociation reactions occur at the target and the films are Si-rich with lowered electrical resistivity. Sputtering in N2 avoids these problems at the expense of deposition rate. Reactive sputtering of a Si target in N2 is used more commonly but is more difficult to control. PECVD nitride deposition has very largely supplanted sputter deposition.

Because of the absence of complex chemical interactions in sputter deposition, the dependencies of rate, uniformity, and film quality on deposition parameters and reactor design are easier to control (and describe).

Step coverage by sputtered films was first modelled by Standley et al (80). The results are similar to those for other films deposited from an extended source, i.e., non-conformality due to shadowing by steps. However, they showed that if the sticking coefficient (1-re-emission coefficient) is low, coverage of the vertical wall at the step is improved, i.e., re-sputtering of the depositing film, due to the substrate bias, redistributes it. Re-deposition on the side walls, rather than almost complete loss, is explained by the fact that at low energies, the angular distribution of sputtered

particles is under-cosine (68d). The effect of substrate bias on step coverage has been illustrated by Kennedy (81) and modelled by others (82).

PLANARIZATION of the INTERLEVEL DIELECTRIC

As each successive layer of a multilevel metal/insulator structure is built, the topography introduced at the contact hole level becomes more severe. Severe topography causes problems of line-width control in in photolithography (used for patterning) as well as in step coverage of the insulator by metal films. A planar structure is desirable, but difficult to realize. Several approaches have been used; some mitigate the effects of severe topography, others attempt to achive complete planarization.

Processes have been developed to modify deposited inorganic insulators by rounding existing steep steps or making the steps shallower so that the next level of metal will cover adequately. These include flowage of P- or P/B- or Ge-doped oxides at high temperature (which was discussed previously), spacer technology, and in-situ modification of PECVD deposition.

As mentioned earlier, polyimides have been used as interlevel dielectrics to planarize partially and smooth underlying steps.

Planarization has been achieved by the fabrication of embedded structures and by the use of a vast variety of etch-back processes for planarizing a deposited insulator. Deposition of planar insulating films has been accomplished by bias sputtering, bias ECR, and by a sequence of deposition/etch cycles for PECVD.

Spacer technology combines deposition of a conformal insulator with maskless anisotropic RIE. The oxide which remains is called a side-wall or spacer; it smooths the edges of the existing steps. Although the shape of the spacer is a complex function of the etching process (83), practical processes have been developed (83b,84).

Embedded structures are built by etching slots into the insulator, depositing metal into the slots by a lift-off process and then filling the gaps between the metal and insulator. In an SiO2 matrix (85) SOG was used to fill gaps; in a polyimide matrix gaps could be filled completely by the appropriate polyimide since the gap was uniform, independent of conductor geometry (18a,86). A planar metal/sputtered-SiO2 structure was formed using MgO as a lift-off layer to remove the excess SiO2 (87); an ECR SiO2 lift-off process used a Mo stencil (88). A multi-step embeddment process, called "PRAIRIE", was described recently (89).

Partial or complete planarization of SiO2 can be accomplished by bias sputtering (90). Re-sputtering of the depositing film is increased by raising the substrate bias substantially; a disadvantage is the decrease in the net accumulation rate. The mechanism proposed for planarization (90a) is based on the fact that the sputter removal rate depends on the angle of incidence of the impinging ions (91). The model assumes that the sputter etch rate at the substrate (the re-sputtering rate) is a function of the angle of the substrate surface, i.e., of the step over which the SiO2 is deposited. This is shown in Fig. 6a, which also shows the deposition rate to be independent of step angle. Whether one makes this assumption or assumes that the side walls receive about half as much material as the horizontal surfaces (80), the general arguments about planarization are still valid. Figure 6b shows a typical metal land to be coated. Reference 90a contains the details of the model that explains the following results. At low re-emission (low re-sputtering, low substrate bias), the profile will be that in Fig.5c. At intermediate values of re-emission, the profile of Fig.5d results; planarization is a burial process. The extent of planarization increases with increasing substrate bias. It is shown as complete in Fig.5d, but it depends on the width of the line to be covered; in practice, wide lines are only partially planarized (Fig.5e). The assumption has been made that the corners of the metal land are not sputtered (which may be approximately valid for Al or Al alloy lands covered with a native oxide that has a very low sputtering rate), but sometimes faceting of the metal lands is observed.

Figure 5f gives the result obtained when the film is deposited at low or intermediate re-emission, followed by an increase in bias to prevent deposition. Since the removal rate on the horizontal surface is lower than that on the slope, the cone of SiO2 will be removed eventually.

This last technique of covering a metal stripe with an insulator and then taking advantage of the enhanced sputter etch rate of a sloped surface has been used to produce tapered PSG and Si3N4 (92a) as well as planar structures in sputter deposited SiO2 (92b,c). A variation, called reactive facet tapering, uses a combination of Ar and CHF3 to taper the edge profile of PECVD SiO2 deposited over metal steps (93).

A process, which may be viewed as analagous to bias sputtering, is bias PECVD deposition, but the objective has been to produce tapered structures. During deposition, one of the reactants also etches the growing film; N2O (94a) and SiCl4, in a process called RID, radical and ion assisted CVD (94b), have been used.

In-situ sequential cycles of PECVD deposition/argon sputter etch (95) were used to deposit conformal or tapered SiO2 in a cold wall LPCVD reactor modified for PECVD. For planar SiO2, a sloped film was deposited to a thickness greater than the underlying metal step and then sputter etched for an extended period.

In order to p'anarize an insulator deposited over closely spaced metal conductors, it is necessary to fill the small spaces without voids. As mentioned earlier, gap-filling becomes increasing difficult as the AR and wall angle increase. Gap-filling, by sputtered SiO2 increases (as does planarization) with increasing substrate bias (i.e., with an increasing re-emission coefficient). Logan (96) has modelled gap-filling. The controlling factor is the sputtering yield vs. angle of incidence curve; the critical value of AR, i.e. the highest AR that can be filled without voids depends on the wall angle and the stable angle which forms due to material deposited on the side wall; the higher the

re-sputtering ratio and the lower the taper angle, the higher the critical AR. It has been shown experimentally that, when the metal side-wall angle is 85° and the sputtering ratio is 0.72, an AR of about 2.3 can be filled with sputtered SiO2 without voids; for vertical walls the critical value of AR drops to about 1.6.

Bias/ECR (97) is a two-step process for gap-filling and planarization; it is illustrated in Fig.6. The normal incidence of the reactant in ECR eliminates shadowing effects so that the deposition rate at the bottom of a small gap is the same as that on the exposed horiziontal surface. By increasing the bias after the oxide thickness equals the metal thickness, the slopes are eroded as the SiO2 is deposited, and a planar surface results. This was illustrated for AR~2 and a wall angle <90° (97b).

One of the earliest methods of planarizing an inorganic dielectric by etching was the use of selective masking (98a). The insulator was coated with resist and a pattern formed in the resist to cover the depressed areas. The elevated regions were then etched. This method has been revisited recently (98b). A related aproach is the use of a dummy pattern (98c). After the interconnects are coated with an insulator, a pattern is formed to cover, approximately, the depressed area. The surface is then coated with a sacrificial layer of polyimide, which is, therfore, planar; the polyimide is then etched back to the required level.

One of the first blanket etch-back processes was based on the use of polyimide as a planarized interlevel dielectric layer (99a). After deposition and curing of a polyimide film, it was dry etched without a mask to expose the interlevel interconnects that had been formed on the conductors. Polyimide was used as the sole insulator in three and five level structures. More recently (99b) a process was developed that used a very thick layer of of a low expansivity polyimide to improve planarity. After curing it was etched without a mask almost to the tops of the underlying metal; a layer of PECVD oxide was deposited and via holes etched in the dual insulator.

There are many etch-back processes based on the use of an inorganic insulator and a spin-coated sacrificial film; the earliest processes used organic films, usually a photoresist, as the sacrificial layer. The simplest concept is shown in Fig.7a. The organic film is spin-coated over an inorganic insulator; it is assumed to be planar. Either an inert or a reactive plasma is used for etching. The sacrificial layer is etched until the the high points of the inorganic film are uncovered; both films are then etched at what is assumed to be the same rate, i.e., ERR = ER (sacrificial layer/ ER (interlevel dielectric) = 1, until the sacrificial layer is completely consumed. Etching can be either stopped within the inorganic layer (Fig.7b), or continued to expose the the underlying metal (Fig.7c). Additional layers of an interlevel dielectric can be deposited, when needed, on the prepared surface and via holes etched. Processes based on this assumption have been described (100).

In the illustrative diagrams, the metal has been shown as planar; this is usually not the case. Therefore, in etching to expose the metal, some surfaces will remain buried, to be reached by via holes. The alternative is to continue etching until the lowest metal surface is exposed. Complete planarity is not attainable.

Moreover, films applied by spin-coating are not planar (Fig.8) although planarization can be improved (but still imperfect) by baking (101). A two-layer resist system for improving planarization has been proposed (102a). The first layer is a low viscosity resist, which partially fills the steps, the second is one of much higher viscosity; after the composite has been formed, it is baked. An analagous approach was the use of low molecular weight (mw) polymers (102b). The dependence of step height on feature size was reduced significanly as mw was decreased from 100,000 to 2,000. However, because the glass transition temperature (Tg) of the low mw polymers is low, an increase in temperature during RIE etch-back degrades the surface smoothness; therefore wafer cooling by heat-sinking to a cooled cathode was required.

The problem of planarizing the interlevel dielectric using a non-planar sacrificial layer has been addressed (103); the solution was varying the ERR.

The other problem is the presumed equality of etch rate. Near-equality of etch rate was demonstrated in a low power, very pure argon plasma (104), but the etch rates were very low. In ion milling, equality was obtained by varying the angle of incidence of Ar+ (105); this too is a slow process. Therefore, reactive ion etching is most commonly used for etch-back. The plasmas contain oxygen to etch the organic film and a fluorinated compound to etch both the Si-containing dielectric and the organic layer. The gas mixture and other RIE parameters such as power, pressure, and flow rate can be varied so that a given etch rate ratio, ERR=ER(insulator)/ER(sacrificial layer), is obtained. However, since RIE has a chemical component, etch rates depend on the area of material to be etched. This is called the loading effect (106). On a large scale, it manifests itself by sensitivity to the number of wafers being etched. What has been called micro-loading is sensitivity to the pattern factor on a given wafer; in this case it would be the relative areas of the two materials. In etch-back, the area ratio changes as the composite is etched and more of the inorganic film is exposed; as a result, the ERR will change although the etching parameters are constant. Therefore, practical etch-back processes must be more complex. Multi-step sequences in which ERR is changed have been devised; these also accomodate the lack of planarity of the sacrificial layer (103).

An alternative to the organic sacrificial layer is SOG; several kinds of materials are available commercially. A solution of the organosilicon polymer is applied to the interlevel dielectric by spin coating and cured. The film fills the grooves and partially planarizes the surface. The SOG layer is etched, either with or without an additional sacrifical resist layer. The SOG is left in the grooves after etch-back, and the structure is then coated with a second layer of a suitable dielectric. Several variations of SOG etch-back have been published (107); one example is shown in Fig. 9.

Deposited inorganic materials have also been used as sacrificial layers. PECVD SiN is an example. In one process, its etch rate in grooves was suppressed compared to its etch rate on flat surfaces, by use of a CF4/H2 plasma (108). In another, (109) a thick film of SiN was deposited over a thin conformal layer of PECVD SiON, the SiN was etched without a mask to the SiON surface, and the resulting near-planar structure coated was coated with SiON as the interlevel dielectric (Fig.10). Still another variation was the combination of a sacrificial glass layer with an organic overcoat; the combination was etched to expose the tops of the metal conductors (109). In addition, there are processes that use several sequences of inorganic dielectric deposition and resist etch-back to form one planar dielectric level (110), as illustrated in Fig.11.

Although drawbacks in etch-back planarization are apparent, it does improve topography significantly and has been used extensively.

PASSIVATING FILMS

Passivation coatings protect the completed semiconductor device structure from the environment, which may be a hostile one. Both hermetic (inorganic) and lower cost polymer-encapsulation are used. The inorganic dielectrics are those used as interlevel insulators. Among the polymers are epoxy resins of various formulations. The restrictions on temperature for deposition of the final passivation overcoating of semiconductor devices are the same as those imposed on deposition of the interlevel dielectrics.

The specific functions of such coatings may include scratch protection of the interconnection metallurgy and resistance to penetration by moisture (protection against corrosion, moisture pick-up by susceptible underlying insulators) and alkali ions. Other requirements are (i) low stress (compressive preferable), (ii) conformal step coverage, (iii) uniformity, (iv) etchability, and (v) good adhesion to underlying metals and insulators.

CONCLUSION

The choice of materials for interlevel dielectrics and passivating films for multilevel devices will be influenced strongly by performance objectives. The methods used for deposition and planarization will probably be selected on the basis of the manufacturing objectives of cost, yield, and reliability.

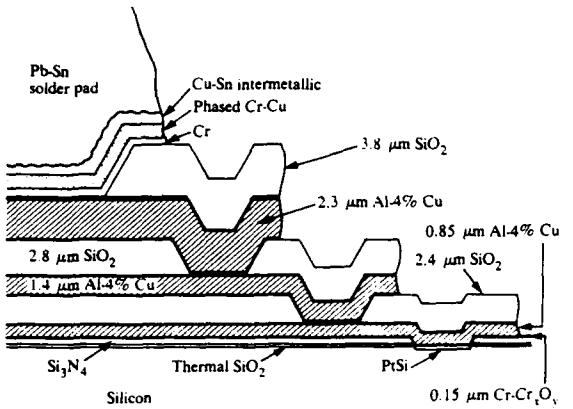


Fig.1 Diagram of a 3 level bipolar structure
From reference 2; reprinted with permission,

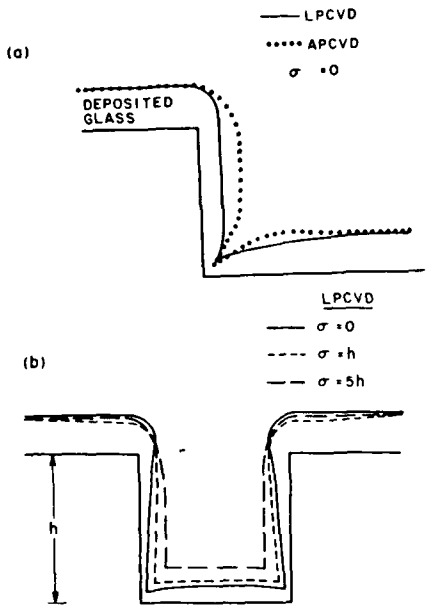


Fig.3 Theoretical profiles of glass films deposited over steps by CVD processes. σ is mean free path for surface migration.
From reference 40; reprinted with permission of the publisher.

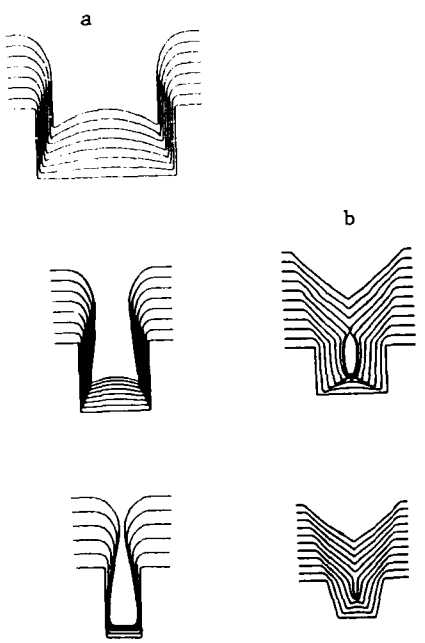


Fig.2 Computer simulation for sputter deposition
(a) Into grooves of different aspect ratios (AR)
From top to bottom: AR = 0.5, 1.0, 2.0
(b) Into vertical and 75- sloped walls (AR = 0.67)
From reference 3; reprinted with permission of the publisher.

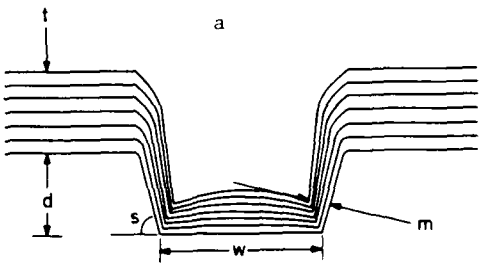


Fig.4(a) Calculated growth contour of PECVD films; t=maximum film thickness, d=step depth, w=interstep width, s=sidewall angle, m=minimum film thickness, m/t=minimum fractional coverage

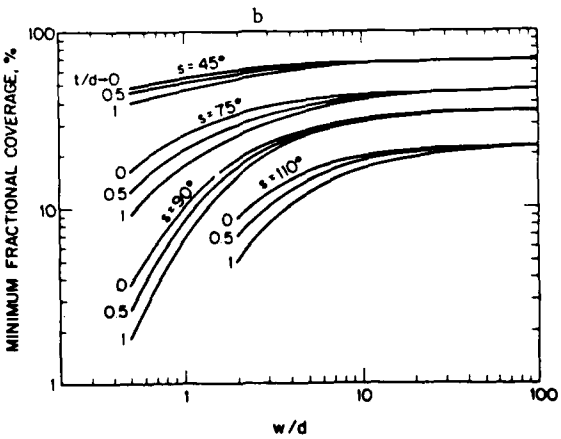


Fig.4(b) m/t vs w/d for various s and t/d
From refernce 66; reprinted with permission of the publisher.

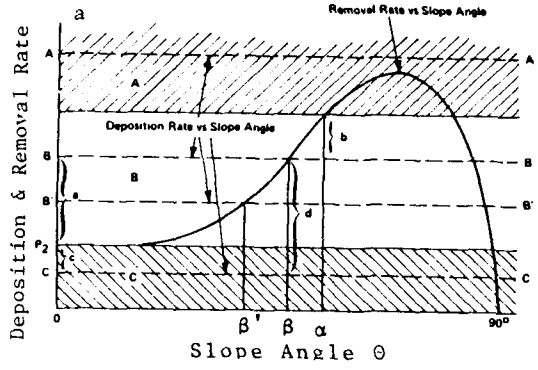


Fig.5(a) Deposition and removal rate vs. slope angle

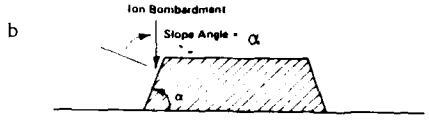


Fig.5(b) Typical shape of a metal land to be coated

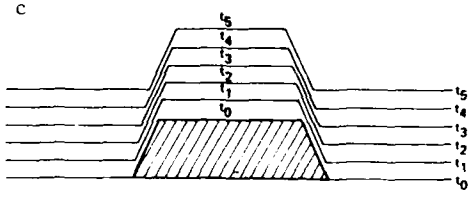


Fig.5(c) Typical non-planarized contours of oxide on metal land

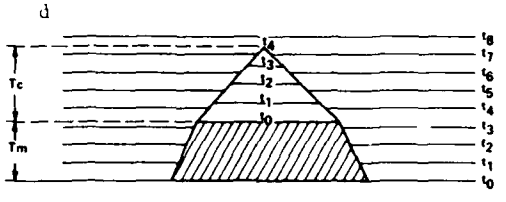


Fig.5(d) Planarized contours of oxide on metal land

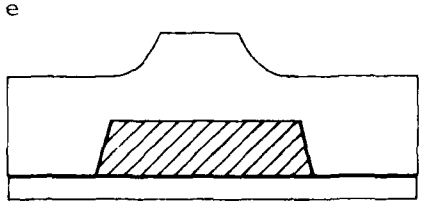


Fig.5(e) Wide land; not completely planarized

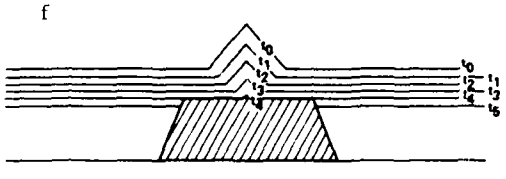


Fig.5(f) Oxide contours during removal result: metal/oxide co-planar

All but Fig.5(e) from reference 90(a); reprinted with permission of the publisher

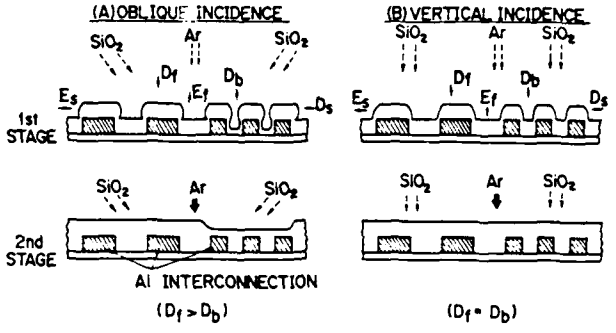


Fig.6 Comparison of the planarization processes with oblique and vertical incidence. Df,Db,Ds are deposition rates on flat, bottom, and side surfaces; E are etch rates. In 1st stage Es<Ds; in 2nd, Es>Ds
From reference 97(a); reprinted with permission of the publisher

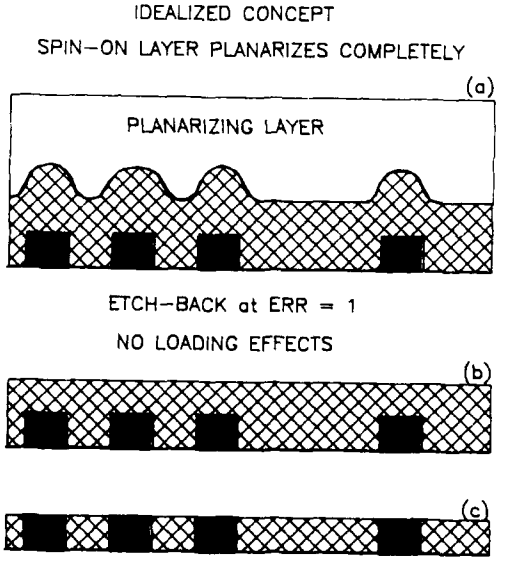


Fig.7(a) Idealized concept of planarization by an organic sacrificial layer

Fig.7(b) After etch-back, stopping in underlying dielectric

Fig.7(c) After etch-back to expose metal lands

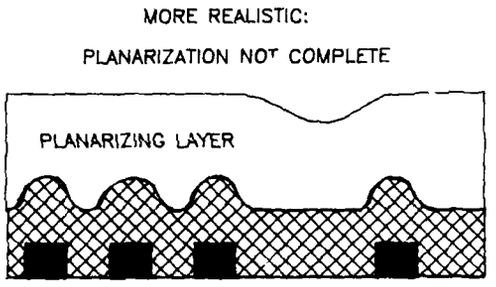


Fig.8 More realistic view of result of coating

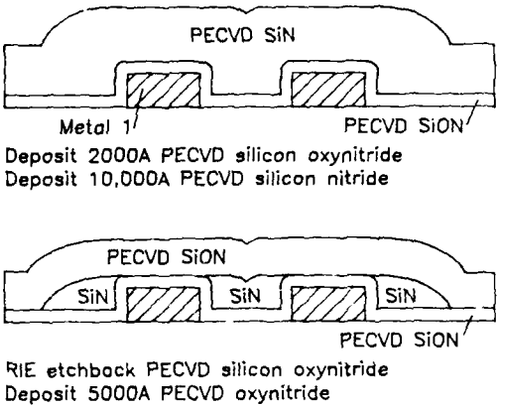


Fig.10 PECVD nitride etch-back sequence
From reference 109; this paper was originally presented at the Fall Meeting of The Electrochemical Society, Inc. held in San Diego, CA

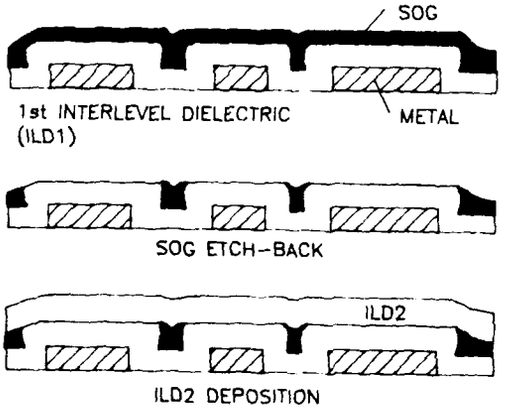


Fig.9 SOG etch-back sequence
From reference 107(a); reprinted with permission
© 1986-IEEE

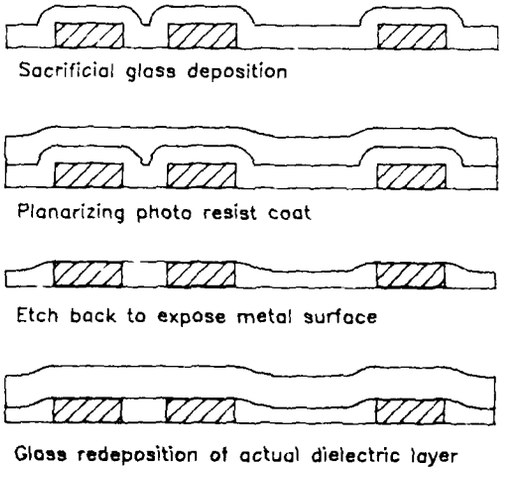


Fig.11 Etch-back planarization process flow using sacrificial glass layer
From reference 110; reprinted with permission
© 1986-IEEE

REFERENCES

(1) (a) E.Berndlmaier, Proc., IEEE Intl. Conf. on Circuits and Computers, 1112 (1980)
(b) S.Brenner, T.A.Bartush, D.J.Swietek, D.C.Banker, F.J.Crispi, J.Deloro, D.L.Merrill, J.P.Norswortny, M.-N.Shen, and C.D.Waggoner, IEEE Intl.Solid-State Circuits Conf., 152 (1983)
(2) L.J.Fried, J.Havas, J.S.Lechaton, J.S.Logan, G.Paal, and P.A.Totta, IBM J.Res.Develop., 26, 363 (1982)
(3) H.P.Bader and M.A.Lardon, J.Vac.Sci.Technol., B-4, 1192 (1986)
(4) (a) J.W.Coburn, Plasma Chemistry and Plasma Processing, 2, 1 (1982)
(b) L.M.Ephrath and G.S.Mathad, in Advanced Semiconductor Technology and Computer Systems, G.Rabbat, ed., Van Nostrand Reinhold, to be published, November 1987
(c) S.J.Fonash,, Solid State Technology, January, 1985, p.150
(d) H.H.Sawin, ibid., April 1985, p.211
(e) R.A.Morgan, Plasma Etching in Semiconductor Fabrication, Elsevier, 1985
(5) (a) J.R.Havas., Electrochem.Soc.Ext.Abstr., 76-2, #289 (1976)
(b) J.M.Moran and D J. Maydan, J.Vac.Sci.Technol., 16, 1620 (1979)
(c) B.J.Lin, Solid State Technology, May 1983, p. 105
(d) E.Ong and E.L.Hu, ibid., June 1984, p. 155
(6) (a) B.L.Cohen, C.L.Fink, and J.H.Degnam, J.Appl.Phys., 43, 19 (1972)
(b) P.Paduschek and P.Eichinger, Appl.Phys.Lett., 36, 62 (1980)
(c) M.Takahashi, M.Maeda, and Y.Sakakibara, Jpn.J.Appl.Phys., 26, 1606 (1987)
(7) (a) D.A.Leich and T.A.Tombrello, Nucl.Instrum.Methods, 108, 19 (1973)
(b) W.A.Lanford, H.P.Trautvetter, J.F.Ziegler, and J.Keller, Appl.Phys.Lett., 28, 566 (1976)
(8) W.A. Lanford and M.J.Rand, J.Appl.Phys., 49, 2473 (1978)
(9) (a) W.A.Pliskin and E.E.Conrad, IBM J.Res.Develop., 8, 42 (1964)
(b) W.A.Pliskin and R.P.Esch, J.Appl.Phys. 36, 2011 (1965)
(10) (a) E.A.Corl and H.Wimpfheimer, Solid State Electronics, 7, 755 (1964)
(b) F.Reizman and W.Van Gelder, ibid., 10, 625 (1967)
(11) (a) P.K.Tien, R.Ulrich, and R.J.Martin, Appl.Phys.Lett., 14, 291 (1969)
(b) A.C.Adams, D.P.Schinke, and C.D.Capio, J.Electrochem.Soc., 126, 1539 (1979)
(12) W.A.Pliskin and R.P.Gnall, ibid., 111, 872 (1964)
(13) (a) W.A.Pliskin and H.S.Lehman, ibid., 112, 1013 (1965)
(b) W.A.Pliskin, J.Vac.Sci.Technol., 14, 1064 (1977)
(c) A.C.Adams, F.B.Alexander, C.D.Capio, and T.E.Smith, J.Electrochem.Soc., 128, 1545 (1981)
(d) A.Takamatsu, M.Shibata, M.Ishida, H.Sakai, and T. Yoshimi, ibid., 133, 443 (1986)
(14) G.Samuelson and S.Lytle, in Polyimides, K. L. Mittal, ed., Plenum Press, NY 1984, p. 751
(15) (a) L.B.Rothman, J.Electrochem.Soc., 127, 2216 (1980)
(b) G.Samuelson, in Polymer Materials for Electronic Applicatins, E.D.Feit and C.W.Wilkins, Jr., eds. ACS Symposium Series 184, American Chemical Society, Washington, D.C., 1982 p.93
(c) A.W.Wilson, D.Laks, and S.M.Davis, ibid., p.139
(16) (a) Y.Misawa, N.Kinjo, M.Hirao, S.Numata, and N.Momma, Proc., 1986 Symposium on VLSI Tech., San Diego, p.59
(b) Y.Misawa, N.Kinjo, M.Hirao, S.Numata, ans N.Momma, IEEE Trans.on Electron Devices, ED-34, 621 (1987)
(17) A.W.Wilson, in Polyimides, K. L. Mittal, ed., Plenum Press, NY 1984, p. 715
(18) Y.K.Lee and J,D.Craig, in Polymer Materials for Electronic Applications, E.D.Feit and C.W.Wilkins, Jr., eds. ACS Symposim Series 184, American Chemical Society, Washington, D.C., 1982, p. 107
(19) J.R.Salem, F.O.Sequeda, J.Duran, W.Y.Lee, and R.M.Yang, J.Vac.Sci.Technol., A-4, 369 (1986)
(20) (a) L.B.Rothman, J.Electrochem.Soc., 130, 1131 (1983)
(b) D.R.Day, D.Ridley, J.Mario, and S.D.Senturia, in Polyimides, K. L. Mittal, ed., Plenum Press, NY 1984, p.767
(c) C.C.Chao and W.V.Wang, ibid., p.783
(21) (a) L.E. Stillwagon, Solid State Technology, June, 1987, p. 67
(b) L.K.White, J.Electrochem.Soc., 132, 168 (1985)
(c) R.H.Wilson and P.A.Piacente, ibid., 133, 981 (1986)
(d) A.Schlitz, P.Abraham, and E.Dechenaux, ibid., 134, 190 (1987)
(22) (a) H.Eggers, H.Fritzsche, and A.Glasl, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 85, 163 (1985)
(b) H.Eggers, A.Glasl, and H. Fritzsche, Siemens Forsch.-u. Entwickl.-Ber. Bd., 15 64 (1986)
(c) R.G.Narechania, J.A.Bruce, and C.A.Beach, Proc., 22nd Reliability Phys. Symp., 214, (1984)
(23) T.Nishida, K.Mukai, T.Inata, I.Tezuka, and N.Horie, Proc., 23rd Reliability Phys. Symp., 148 (1985)
(24) A.M.Wilson, Electrochem.Soc.Ext.Abstr., 80-2, #495 (1980)
(25) (a) C.Mitchell and R.Goodner, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 84, 130 (1984)
(b) H.L.Peek and R.A.M.Wolters, ibid., V-MIC 86, 165 (1986)
(26) D.R.Day, R.Whitten, and S.D.Senturia, Electrochem.Soc.Ext. Abstr., 81-2, #242 (1981)

(27) P.deGeyter, G.Brasseur, and F.Coopmans, in Multilevel Metallization, Interconnection and Contact Technologies, L.B.Rothman and T.Herndon, eds., The Electrochem.Soc., PV 87-4, 67 (1987)
(28) G.A.Brown, in Polymer Materials for Electronic Applications, ACS Symposium Series 184, American Chemical Society, Washington, D.C., 1982, p.151
(29) (a) J.R.Lloyd, Thin Solid Films, 91, 175 (1982)
(b) J.R.Lloyd and R.N.Steagall, J.Appl.Phys., 60, 1235 (1986)
(30) S.S.Iyer and C.-Y.Ting, IEEE Trans.Electron Devices, ED-31, 1468 (1984)
(31) (a) I.V.Rivas, K.S.Ravindhran, and N.M.Kulkarni, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 84, 283 (1984)
(b) A.Saiki, S.Harada, T.Okubo, K.Mukai, and T.Kimura, J.Electrochem.Soc., 124, 1619 (1977)
(c) S.J.Rhodes, Semiconductor International, March, 1981, p.65
(d) D.L.Bergeron, J.F.Kent, and K.E.Morrett, Proc., 22nd Reliability Phys. Symp., 1984, p.229
(e) H.Tsunetsuga, A.Takagi, and K.Moriya, Intl. J. for Hybrid Electronics, 8 21 (1985)
(f) J.J.Lajza, M.A.Alfonso, J.T.Weed, and J.L.Wendt, in Multilevel Metallization, Interconnection and Contact Technologies, L.B.Rothman and T.Herndon, eds., The Electrochem.Soc., V 87-4, 250 (1987)
(32) (a) G.C.Davis, in Polymers in Electronics, T. Davidson, ed., ACS Symposium Series 242, American Chemical Society, Washington, D.C., 1984, p. 259
(b) N.Kabuhiko, T.Wada, N.Naito, and N.Mutoh, IEEE Electron Device Lett., EDL-6, 589 (1985)
(33) (a) W.A.Pliskin and P.P.Castrucci, Electrochem.Technol., 6, 85 (1968)
(b) H.Sankur, J.G.Nelson, A.T.Pritt,Jr., and W.J.Gunning, J.Vac. Sci.Technol., A-5, 15 (1987)
(c) J.Wong, T.-M.Lu, and S.Mehta, ibid., B-3, 453 (1985)
(f) J.M.E.Harper, in Thin Film Processes, J.L.Vossen and W.Kern, eds., Academic Press, 1978, p.175
(g) Y.Minowa and H.Ito, J.Vac Sci.Technol., B-6, 473 (1987)
(h) R.Solanki, C.A.Moore, and G.J.Collins, Solid State Technology, June, 1985, p. 220
(i) J.Shirafuji, S.Miyoshi, and H.Aoki, Thin Solid Films, 157, 105 (1988)
(j) K.Inoue, M.Michimori, M.Okuyama, and Y.Hamakawa, Jpn.J.Appl. Phys., 26, 805 (1987)
(k) J.Marks and R.E.Robinson, Appl.Phys.Lett., 52, 810 (1988)
(l) M.Kuisl, Thin Solid Films, 157, 129 (1988)
(m) A.D.Butherus, T.W.Hou, C.J.Mogab, and H.Schonhorn, J.Vac.Sci. Technol., B-3, 1352 (1985)
(34) (a) H.Tsunemitsu and H.Shiba, IEEE International Electron Devices Meeting, October 1969, Abstr. #18.4
(b) W.R.McMahon, ibid., October, 1970, Abstr. #7.1
(c) G.C.Schwartz and V.Platter, J.Electrochem.Soc., 122, 1508 (1975
(d) G.C.Schwartz and V.Platter, ibid., 123, 34 (1976)
(e) H.Tsunemitsu and H.Shiba, Semiconductor Silicon, 1977, H.R.Huff and E.Sirtl, eds., The Electrochemical Society, PV 77-2, 1035 (1977)
(35) (a) J.R.Franco, J.Havas, J.Rompala, US Pat 4,004,044 (1977)
(b) T.Sakurai and T.Serikawa, J.Electrochem.Soc., 126, 1257 (1979)
(c) J.Havas, US Pat 4,202,914 (1980)
(d) M.Hatzakis, B.J.Canavello, and J.M.Shaw, IBM J.Res.Develop., 24, 452 (1980)
(e) J.L.Speidell, J.Vac.Sci.Technol., 19, 693 (1981)
(f) Y.Homma, H.Nozawa, amd S.Harada, IEEE Trans. on Electron Devices, ED-28, 552 (1981)
(g) P.Grabbe, E.L.Hu, and R.E.Howard, J.Vac.Sci.Technol., 21, 33 (1982)
(h) Y.Homma, A.Yajima, and S.Harada, IEEE J. Solid-State Circuits, SC-17, 142 (1982)
(i) L.D.Hou, Electrochem.Soc.Ext.Abstr., 83-2, #213 (1983)
(j) J.B.Dinklage and M.C.Hakey, IBM Tech.Disclosure Bull., 27, 2800 (1984)
(k) E.H.Sebesta, US Pat 4,497,684 (1985)
(l) H. Moritz, IEEE Trans.on Electron Devices, ED-32, 672 (1985)
(m) P.Pai, Y.Shacham-Diamand, and W.G.Oldham, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 86, 209 (1986)
(n) H. Shibata, US Pat 4,576,678 (1986)
(o) H.S.Sachdev and K.G.Sachdev, US Pat 4,599,243 (1986)
(36) (a) D.W.Hess, Plasma Chemistry and Plasma Processing, 2, 141 (1982)
(b) G.C.Schwartz, in Plasma Processing, G.S.Mathad, G.C.Schwartz and G.Smolinsky, eds., The Electrochemical Society, PV 85-1 26, (1985)
(37) (a) W.Kern and V.S.Ban, in Thin Film Processes, J.L.Vossen and W.Kern, eds., Academic Press, 1978, p.257
(b) J.-O.Carlsson, Thin Solid Films, 130, 261 (1985)
(38) (a) W.Kern and R.S.Rossler, J.Vac.Sci.Technol., 14, 1082 (1977)
(b) J.A.Amick, G.L.Schnable, and J.L.Vossen, ibid., p.1053
(c) B.Gorowitz, R.H.Wilson, and T.B.Gorczyca, Solid State Technology, Oct. 1987, p. 97
(c) Chemical Vapor Deposition, 1960-1980: A Bibliography D.T.Hawkins, ed., IFI/Plenum Data Co., New York (1981)
(d) D.W.Hess, K.F.Jensen, and T.J.Anderson, Rev.Chem.Eng., 3, 97 (1985)
(e) D.R.Biswas, J.Mater.Sci., 21, 2217 (1986)
(39) (a) A.J.Learn and R.B.Jackson, J.Electrochem.Soc., 132, 2975 (1985)
(b) J.A.Aboaf, ibid., 116, 1732 (1969)

(c) A.C.Adams and C.D.Capio, ibid., 127, 1042 (1979)
(d) K.Maeda and J.Sato, Denki Kagaku, 45, 645 (1977)
(40) R.M.Levin and K.Evans Lutterodt, J.Vac.Sci.Technol., B-1, 54 (1983)
(41) (a) W.Kern and R.K.Smeltzer, Solid State Technology, June, 1985,
(b) R.M.Levin and A.C.Adams, J.Electrochem.Soc., 129, 1588 (1981)
(c) A.C.Adams, C.D.Capio, S.E.Haszko, G.I.Parisi, E.I.Povilonis, and McD. Robinson, ibid., 126, 313 (1979)
(d) F.C.Chien, R.L.Brown, G.N.Burton, and M.B.Vora, IEEE VLSI Multi level Interconnection Conf., V-MIC 84, 45 (1984)
(42) (a) H.J.Stein and H.A.R.Wegener, J.Electrochem.Soc., 124, 908 (1977
(b) P.S.Peercy, H.J.Stein, B.L.Doyle, and S.T.Picraux, J.Electron. Mat., 8, 11 (1979)
(43) D.W.Hess, J.Vac.Sci.Technol., A-2, 244 (1094)
(44) W.A.P.Claassen, Plasma Chemistry and Plasma Processing, 7, 109 (1987)
(45) (a) S.W.Ing, Jr., and W. Davern, J.Electrochem.Soc., 112, 284 (1965)
(b) D.R.Secrist and J.D.Mackenzie, ibid., 914, (1966)
(c) R.J.Joyce, H.F.Sterling. and J.H.Alexander, Thin Solid Films, 1, 481 (1967/68)
(46) (a) A.R.Reinberg, Electrochem.Soc.Ext.Abstr., 74-1, #6 (1974)
(b) A.R.Reinberg, J.Electron.Mat., 8, 345 (1979)
(47) (a) J.R.Hollahan and R.S.Rosler, in Thin Film Processes, J.L.Vossen and W.Kern, eds., Academic Press, 1978, p.335
(b) M.J.Rand, J.Vac.Sci.Technol., 16, 420 (1979)
(c) A.R.Reinberg, J.Electron.Mat., 8, 345 (1979)
(d) A.C.Adams, Solid State Technology, April 1983, p.135
(e) D.W.Hess, J.Vac.Sci.Technol., A-2, 244 (1984)
(f) A.C.Adams, in Plasma Processing, R.A.Gottscho, G.S.Mathad, and G.C.Schwartz, eds., The Electrochemical Society, Pennington, NJ, PV 87-6, 375 (1987)
(g) B.Gorowitz, R.H.Wilson, and T.B.Gorczyca, Solid State Technology, Oct. 1987, p. 97
(48) G.Kaganowicz, V.S.Ban, and J.W.Robinson, J.Vac.Sci.Technol., A-2, 1233, 1984
(49) (a) R.S.Rosler and G.M.Engle, Solid State Technology, December, 1979, p.88
(b) R.S.Rosler and G.M.Engle, ibid., April, 1981, p. 172
(50) K.Law, J.Wong., and D.N.K.Wang, Technical Proceedings, Semicon Japan, 1987, p. 154
(51) E.P.van de Ven, R.S.Martin, and M.J.Berman, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 87, 434 (1987)
(52) P.G.Pai, S.S.Chao, Y.Takagi, and G.Lucovsky, J.Vac.Sci.Technol., A-4, 689 (1986)
(53) (a) M.J.Helix,K.V.Vaidyanathan, B.G.Streetman, H.B.Dietrich, and P. K. Chaterjee, Thin Solid, Films, 55, 143 (1978)
(b) L.G.Meiners, J.Vac.Sci.Technol., 21, 655 (1982)
(c) K.P.Pande and P.W.Davies, J.Electron.Mat., 13, 593 (1984)
(d) P.D.Richard, R.J.Markunas, G.Lucovsky, G.G.Fountain, A.N.Mansour, and D.V.Tsu, J.Vac.Sci.Technol., A-3, 867 (1985)
(e) D.V.Tsu and G.Lucovsky, ibid., A-4, 480 (1986)
(f) G.Lucovsky, P.D.Richards, D.V.Tsu, S.Y.Lin, and R.J.Markunas, ibid., p.681
(g) G.Lucovsky, P.D.Richard, and D.V.Tsu, Mat.Res.Soc.Symp.Proc. Vol., 54, 529 (1986)
(h) D.V.Tsu, G.Lucovsky, and M.J.Mantini, Phys.Rev. B, 33, 7069 (1986)
(54) (a) J.E.Spencer, R.L.Jackson, and A.Hoff, in Plasma Processing, R.A.Gottscho, G.S.Mathad, and G.C.Schwartz, eds., The Electrochem.Soc., PV 87-6, 186 (1987)
(b) R.L.Jackson, J.E.Spencer, J.L.McGuire, and A.M.Hoff, Solid State Technology, April 1987, p. 107
(55) (a) S.Matsuo and M.Kiuchi, Jpn.J.Appl.Phys., 22, L210 (1983)
(b) M.Doki, K.Takasaki, and Y.Ban, Electrochem.Soc.Ext.Abstr., 87-1 #254 (1987)
(c) T.Hirao, K.Setsune, M.Kitagawa, T.Kamada, K.Wasa, K.Tsukamoto, and T.Izumi, Jpn.J.Appl.Phys., 27, 30 (1988)
(d) T.Hirao, K.Setsune, M.Kitagawa, T.Kamada, K.Wasa, and T.Izumi, ibid., 26, 2015 (1987)
(56) K.Machida and H.Oikawa, Ext.Abstr.17th Conf. on Solid State Devices and Materials, Tokyo, 1985, p. 329
(57) K.Yamazaki, Y.Numasawa, K.Hamano, T.Mizuno, and J.H.Coleman, Electrochem.Soc.Ext.Abstr., 85-2, #260 (1985)
(58) H.Mito and A.Sekiguchi, J.Vac.Sci.Technol., A-4, 465 (1986)
(59) J.Kanicki and P.Wagner, in Silicon Nitride and Silicon Dioxide Thin Insulating Films, V.J.Kapoor and K.T.Hankins, eds., The Electrochem.Soc., PV 87-10, 261 (1987)
(60) (a) R.B.Fair and R.C.Sun, IEEE Trans. on Electron Devices, ED-28, 8 (1981)
(b) R.C.Sun, J.T.Clemens, amd J.T.Nelson, Proc., 18th IEEE Reliability Phys.Symp., 244 (1980)
(c) V.S.Dharmadhikari, Thin Solid Films, 153, 459 (1987)
(d) R.C.Budhani, R.F.Bunshah, and P.A.Finn, Appl.Phys.Lett., 52, 284 (1988)
(61) A.C.Adams, F.B.Alexander, C.D.Capio, and T.E.Smith, J.Electrochem.Soc., 128, 1545 (1981)
(62) P.Pan, L.A.Nesbit, R.W.Douse, and R.T.Gleason, ibid., 132, 2012 (1985)
(63) (a) S.P.Mukherjee and P.E.Evans, Thin Solid Films, 14, 105 (1972)
(b) L.J.Olmer, E.R.Lory, W.W.Clauhs, and A.Harrus, Electrochem.Soc. Ext.Abstr., 87-1, #255 (1987)

(c) B.L.Chin, E.P.Van de Ven, and S.C.Avanzino, Electrochem.Soc. Abstr.Recent News Papers, #1826 RNP, J.Electrochem.Soc., 134, 624C (1987)
(64) (a) V.S.Nguyen, S.Burton, and P.Pan, ibid., 131, 2348 (1984)
(b) C. Cavallari and F. Gualandris, ibid., 134, 1265 (1987)
(c) T.Hairo, K.Setsune, M.Kitagawa, T.Kamada, T.Ohmura, K.Wasa, and T.Izumi, Jpn.J.Appl.Phys., 27, L21 (1988)
(65) J.A.Thornton, Thin Solid Films, 107, 3, (1983)
(66) R.C.Ross and J.L.Vossen, Appl.Phys.Lett., 45, 239 (1984)
(67) (a) G.K.Wehner and G.S.Anderson, in Handbook of Thin Film Technology, L.I.Maissel and R. Glang, eds., McGraw-Hill, Inc., 1970, p.3-1
(b) L.Maissel, ibid., p.4-1
(c) J.L.Vossen and J.J.Cuomo, in Thin Film Processes, J.L.Vossen and W. Kern, eds., Academic Press, 1978, p.12
(d) Brian Chapman, Glow Discharge Processes, John Wiley & Sons, 1980
(68) J.W.Coburn, E.Taglauer, and E.Kay, Proc. 6th Internatl.Vacuum Cong., 1974, p. 501
(69) J.H.Keller, and R.G.Simmons, R.G., IBM J.Res.Develop., 23, 24 (1979)
(70) P.Grosewald, L.V.Gregor, L.V and R. Powlus, Proc. IEEE Electron Device Meeting, Washington, DC, Oct 1971
(71) (a) J.S.Logan, J.M.Keller, and R.G.Simmons, J.Vac.Sci.Technol., 14, 92 (1977)
(b) Y.Hazuki and T.Moriya, IEEE Trans. on Electron Devices, ED-34, 628 (1987)
(72) (a) E.Holland and D.S.Campbell, J.Mat.Sci., 3, 544 (1968)
(b) R.M.Valletta, J.A.Perri, and J.Riseman, Electrochem.Technol., 4, 402 (1966)
(c) S.M.Hu amd L.V.Gregor, J.Electrochem.Soc., 114, 826 (1967)
(d) W.Pasadowski, Thin Solid Films, 69, 149 (1980)
(e) P.S.McLeod, Solid State Technology, Oct. 1983, p. 207
(73) J.S.Logan, F.Jones, J.Costable, and J.E.Lucy, J.Vac.Sci.Technol., A-5, 1879 (1987)
(74) (a) R.E.Jones, M.A.Zuegel, and G.C.Schwartz. Proc. 4th Internatl. Vacuum Cong., 1968, p. 557
(b) G.C.Schwartz and R.E.Jones, IBM J.Res.Develop., 14, 52 (1970)
(75) H.R.Koenig, and L.I.Maissel,ibid., p.168
(76) (a) J.L.Vossen, J.Vac.Sci.Technol., 8, S12 (1971)
(b) J.S.Logan, IBM J.Res.Develop., 14, 172 (1970)
(c) A.W.Stephens, J.L.Vossen, and W.Kern, J.Electrochem.Soc., 123, 303 (1976)
(77) (a) R.E.Jones, C.L.Standley, and L.I.Maissel, J.Appl.Phys., 38, 4656 (1967)
(b) L.I.Maissel, R.E Jones, and C.L.Standley, IBM J.Res.Develop., 14, 176 (1970)
(78) (a) J.S.Logan,ibid., p.172
(b) R.Auyang, and K.Jaeckel, IBM, East Fishkill Facility, Hopewell Junction, NY, unpublished results
(79) (a) J.A.Thornton and A.S.Penfold, in Thin Film Processes, J.L.Vossen and W.Kern, eds., Academic Press, 1978, p.76
(b) D.B.Fraser, ibid., p.115
(c) R.K.Waits, ibid., p.131
(80) C.L.Standley, R.E.Jones, and L.I. Maissel, Thin Solid Films, 5, 355 (1970)
(81) T.N.Kennedy, J.Vac.Sci.Technol., 13, 1135 (1976)
(82) (a) A.R.Neureuther, C.H.Ting, and C.-Y.Liu. IEEE Trans.on Electron Devices, ED-27, 1449 (1980)
(b) T.Serikawa, J.Vac.Sci.Technol., 17, 582 (1980)
(83) (a) P.J. Tsang, J.F.Shepard, S.Ogura, and J, Riseman, Electrochem.Soc.Ext.Abstr., 82-2, #233 (1982)
(b) S.J.H.Brader, J.Rogers, and S.C.Quinlan, IEEE VLSI Multilevel Interconnection Conf., V-MIC 86, 93 (1986)
(84) (a) M.Khan, Y.Strunk, and S.Huang, ibid., V-MIC 85, 32 (1985)
(b) W.W.Yao, I.W.Wu, R.T.Fulks, and H.A.Vander Plas, ibid., p.38
(85) Y.Sumitomo and Y.Ohashi, US Pat 4,123,565 (1978)
(86) L.B.Zielinski, US Pat 3,985,597 (1976)
(87) L.B.Rothman, P.M.Schaible, and G.C.Schwartz, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 85, 131 (1985)
(88) K.Ehara, T.Morimoto, S.Muramoto, T.Hosoya, and S.Matsuo, 1982 Symp. on VLSI Technol., Digest ofPapers, Jpn.Soc.Appl.Phys., IEEE, Oiso, Japan, Sept. 1982, #2-7, p.30
(89) L.Wu, in Multilevel Metallization, Interconnection and Contact Technologies, L.B.Rothman and T.Herndon, eds., The Electrochem.Soc., PV 87-4, 239 (1987)
(90)(a) C.Y.Ting, V.J.Vivalda, and H.G.Schaefer, J.Vac.Sci.Technol., 15, 1105 (1978)
(b) S.Tsunekawa, Y.Homma, and S.Harada, Electrochem.Soc.Ext.Abstr., 83-2 #282 (1983)
(c) R.C.Mosely, C.A.Sans, S.D.Campbell, and L.T.Lamont,Jr., J.Vac.Sci.Technol., A-3, 596 (1985)
(91) A.D.Stewart and M.W.Thompson, J.Mat.Sci., 4, 56 (1969)
(92) (a) H.Kotani, H.Yakushiji, H.Harada, K.Tsukamoto, and T.Mishioka, J.Electrochem.Soc., 130, 645 (1983)
(b) T.Mogami, M.Morimoto, H.Okabayashi, and E.Nagasawa, J.Vac.Sci.Technol., B-3, 857 (1985)
(c) B.Singh, O.Mesker, and D.Devlin, ibid., B5, 567 (1987)
(93) T.Abraham, IEEE VLSI Multilevel Interconnection Conf., V-MIC 87 115 (1987)
(94) (a) C.G.Smith and A.J.Purdes, J.Electrochem.Soc., 132, 2721 (1985)
(b) M.Sato and Y.Arita, Jpn.J.Appl.Phys., 25, L764 (1986)

(95) (a) E.J.McInerney, Proc. IEEE VLSI Multilevel Interconnection Conf. V-MIC 86, 467 (1986)
(b) E.J.McInerney and S.C.Avanzino, IEEE Trans.on Electron Devices, ED-34, 615 (1987)
(96) J.S.Logan, IBM, T.J.Watson Research Center, Yorktown Heights, NY, unpublished results
(97) (a) K.Machida and H.Oikawa, J.Vac.Sci.Technol., B-4, 818 (1986)
(b) K.Machida and H.Oikawa, Symp. on VLSI Technol., 1987, p. 69
(98) (a) B.-C.Feng, US Pat 3.976,524 (1976)
(b) A.Schlitz and M.Pons, J.Electrochem.Soc., 133, 178 (1986)
(c) K.Mitsuhashi, SEMI Japan'86, Paper E1-1
(99) (a) K.Sato, S.Harada, A.Saiki, T.Kimura, T.Okubo, and K.Mukai, IEEE Trans. on Parts, Hybrids, and Packaging, PHP-9 176 (1973)
(b) Y.Misawa, N.Kinjo, M.Hirao, and N.Momma, IEEE Trans. on Electron Devices, ED-34, 621 (1987)
(102)(a) R.J.Patterson and M.L.Lust, IEEE ibid., VLSI Multilevel Interconnection Conf., V-MIC 84, 51 (1984)
(b) T.A.Bartush, ibid., V-MIC 87, 41 (1987)
(101 (a) A.C.Adams and C.D.Capio, J.Electrochem.Soc., 128, 423 (1981)
(b) T.R.Pampalone, J.J,DiPiazza, D.P.Kanen, ibid., 133, 2394 (1986
(102)(a) S.Crapella and F.Gualandris, ibid., 135, 683 (1988)
(b) H.Gokan, M.Mukainaru, and N.Endo, ibid., p. 1019
(103)(a) G.E.Gimpelson and C.L.Russo, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 84, 37(1984)
(b) S.S.Tai and F.Choi, ibid., V-MIC 85, 138 (1985)
(104) Y.Hom-ma, S.Harada, and T.Kaji, J.Electrochem.Soc., 126, 1531 (1979)
(105)(a) T.Kubo, US Pat 4,614,563 (1986)
(b) L.F.Johnson, K.A.Ingersoll, and J.V.Dalton, J.Vac.Sci.Technol., B-1, 487 (1983)
(106) G.C.Schwartz, L.B.Zielinski, and T.Schopen, in Etching for Pattern Definition, H. G. Hughes and M. J. Rand, eds., The Electrochem. Soc., 1976, p.122
(107)(a) P.Elkins, K.Reinhardt, and R.Tang., IEEE VLSI Multilevel Interconnection Conf., V-MIC 86, 100 (1986).
(b) J.K.Chu, J.S.Multani, S.K.Mittal, J.T.Orton, and R.Jecmen, ibid., p.474
(c) A.Rey, D.LaFond, J.M.Mirabel, M.C.Tacussel, and M.F.Coster, ibid., p.491
(d) L.B.Vines and S.K.Gupta, ibid., p.506
(e) A.Schlitz, Microelectronic Engr., 5 413 (1986)
(f) P.E.Riley and E.D.Castel , in Multilevel Metallization, Interconnection and Contact Technologies, L.B.Rothman and T. Herndon, eds., The Electrochem.Soc., PV-87-4, 194 (1987)
(g) N.Parekh, R.Allen, W.Yao, and R.Fulks, ibid., p.221
(h) C.H.Ting, H.Y.Lin, P.L.Pai, and W.G.Oldham, ibid., V-MIC 87, 61 (1987)
(i) R.M.Brewer and R.A.Gasser, Jr., ibid., p.376
(108) Y.Hazuki, T.Moriya, and M.Kashiwagi, 1982 Symp., VLSI Tech., Digest of Papers, Japan.Soc.Appl.Phys., IEEE Paper 2-1, p.18
(109) J.L.Sum, G.W.Ray, S.-H Hsu, P.J.Marcoux, J.Kruger, C.-H.Lin, and S.Peng, in Multilevel Metallization, Interconnection, and Contact Technologies, L.B.Rothman and T.Herndon, eds., The Electrochem. Soc., PV 87-4,259 (1987)
(110) W.Geiger and A.Sharma, Proc., IEEE VLSI Multilevel Interconnection Conf., V-MIC 86, 128 (1986)
(111)(a) L.Koyama and M.Thomas, ibid., V-MIC 85, 45 (1985)
(b) H.Fritzche, V.Grewal, and W.Henkel, ibid., V-MIC 86, 45 (1986)
(c) T.D.Bonifeld, R.J.Gale, B.W.Shen, G.C.Smith, and C.H.Huffman, ibid., p.71

CHEMICAL VAPOR DEPOSITION OF DIELECTRIC AND METAL FILMS

Dennis W. Hess
Dept. of Chemical Eng., Univ. of California
Berkeley, California 94720, USA

ABSTRACT

Thin film materials are used in many technologically important areas such as microelectronic devices, optical devices, magnetic devices, solar energy conversion, decoration, passivation, and wear resistance. Numerous thin film formation methods exist. However, due to the versatility and throughout capability of chemical vapor deposition (CVD), this technique has been used to deposit an extensive variety of thin film materials. Control of CVD processes is dominated by fundamental chemical reaction and physical transport principles. Proper implementation of these principles permits the deposition of dielectric and metallic films with specific properties

THE PURPOSE of this paper is to describe the fundamental chemical and physical processes involved in chemical vapor deposition (CVD) of thin film materials, and to discuss the application of this film formation technique to dielectric and metallic films. Since the primary use of CVD is in the fabrication of electronic and optical devices, emphasis will be placed on films employed widely in these technological arenas.

The term chemical vapor deposition is quite descriptive. This process utilizes chemically reactive vapors to synthesize or deposit a film or coating. Various reaction schemes such as pyrolysis, disproportionation, reduction, oxidation, and hydrolysis have been utilized. The energy needed to promote chemical bond breaking is usually supplied by thermal means, but photons or discharges can also be invoked. A CVD system is therefore a chemical reactor. Thus, flow rates and flow patterns of reactant vapors, along with substrate temperature must be precisely controlled if uniform film layers are to be obtained. In addition, prediction of deposition rates and thickness and composition uniformity requires a detailed understanding of thermodynamics, kinetics, fluid flow and mass transport phenomena for the appropriate reactions and reactor designs. Finally, control of film structure and properties demands a knowledge of nucleation phenomena and impurity and defect incorporation during growth.

The objectives of CVD are to form film materials with reproducible and controllable properties at reasonable deposition rates and costs while minimizing the effect of deposition on the results of previous processing. Important film properties include film thickness, composition, purity, crystalline perfection, and surface morphology. Clearly, the variation permitted for each of these properties depends upon the application. For a specific deposition chemistry and reactor design, the process variables available are molar flow rates, energy input, system pressure, and substrate condition. Ideally, one would like to be able to predict CVD results as a function of process variables so that an optimum process can be obtained. Unfortunately, the relationship between process variables and process results is generally complicated and thus poorly understood. Most of the efforts to data have therefore been "trial-and-error" approaches to obtaining specific film properties. Only within the last 5-10 years has significant attention been paid to the fundamental chemistry and reactor design. The basic concerns, approaches, and results of CVD processes are the subject of this review.

FUNDAMENTAL CHEMICAL AND PHYSICAL PRINCIPLES OF CVD

The sequence of steps involved in a CVD process is depicted in Fig. 1. Thermal decomposition and/or reaction of chemical

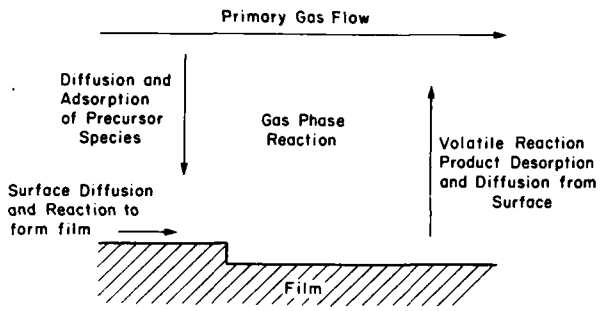


Fig. 1. Steps involved in CVD processes.

species introduced into the reactor occurs in the main gas flow to form reactive intermediates or precursor molecules. These molecules diffuse to the substrate surface where they adsorb, and undergo surface diffusion and reaction to form film material and volatile byproducts. Volatile Products desorb from the surface and diffuse into the primary gas stream whereupon they are swept from the reactor. The overall rate of film deposition is determined by the rate of the slowest step in this sequence. Chemical reactions occur both in the gas phase (homogeneous reactions) or on the growing film surface (heterogeneous reactions). Homogeneous nucleation is to be avoided in order to minimize particulate formation and thus induce heterogeneous reactions which result in the production of uniform continuous films.

Since a basic understanding of CVD involves thermodynamic, kinetic, and transport phenomena, these considerations are now discussed individually. For additional details, see References [1-5].

THERMODYNAMICS - In CVD, a non-equilibrium initial (inlet) state approaches an equilibrium state by chemical reaction and mass transfer. The chemical reaction and phase equilibria relationships provide a description of the reactor performance possible. That is, equilibrium analysis is useful to predict the feasibility of processes, discriminate among reaction mechanisms, and predict gas and solid phase equilibria. However, since CVD is inherently a non-equilibrium process controlled by kinetics and transport phenomena, it is seldom possible to predict quantitative results such as deposition rates.

An equilibrium model for CVD systems involves a gas phase (usually treated as an ideal gas mixture) in contact with pure condensed phases (sources, substrates, reactor materials) and at times, condensed solutions (e.g., multicomponent sources). Computation of the chemical equilibrium state in a complex system is generally formulated in one of two ways [7,8]. In the nonstoichiometric approach, compositions are computed at constant temperature and pressure by minimizing the Gibbs Free Energy of the system

$$\Delta G = \sum_{n=1}^{N} n_i \mu \qquad (1)$$

where N is the number of chemical species, and n_i is the number of species i with chemical potential μ_i. Minimization of ΔG is subject to elemental abundance and mole number non-negativity constraints

$$b_j = \sum_{i=1}^{N} a_{ij} n_i \quad , \quad j = 1, 2, \ldots N \qquad (2a)$$

$$n_i \geq 0 \quad , \quad i = 1, 2, \ldots N \qquad (2b)$$

where a_{ij} is the number of atoms of element j in species i, and b_j is the number of atoms of element j.

In the stoichiometric approach, the species mole numbers are linearly transformed into a new set of molar extent of reaction variables and the system Gibbs Free Energy is minimized in terms of these new variables. The transformation involves the stoichiometric coefficients of an independent set of reactions that describe the chemical system. This approach appears to be well-suited for the analysis of CVD systems.

KINETICS - Both gas phase and surface reactions are complex in CVD; many kinetic studies have therefore been phenomenological. CVD kinetic data have traditionally been reported in terms of the logarithm of the growth (deposition) rate versus substrate temperature (Arrhenius plot) as shown in Fig. 2. At relatively low temperature,

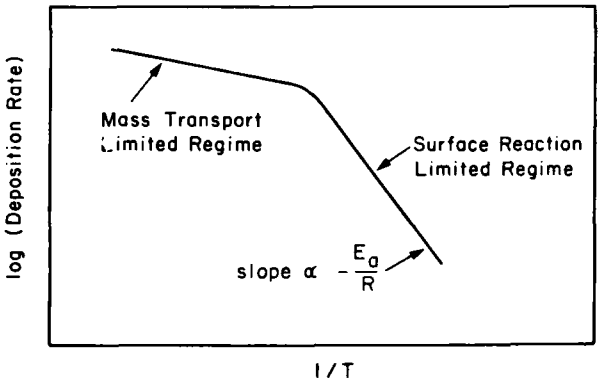


Fig. 2. Temperature dependence of CVD growth rate.

chemical reactions are slow relative to transport phenomena so that deposition rates are thermally activated and obey an Arrhenius expression

$$k = k_o \exp(-E_a/RT) \quad (3)$$

where k is the rate constant, k_o is a pre-exponential factor, E_a is the apparent activation energy of the reaction, and R is the ideal gas constant. As the substrate temperature increases, surface reaction rates increase substantially and growth rates become limited by transport of reactants to the surface, which is only weakly temperature dependent. Unfortunately, these data are only marginally useful in reactor analysis and design. In principle, Langmuir-Hinshelwood or Langmuir-Rideal rate expressions can be formulated in a similar manner to that for catalytic reactions [9,10]. Such rate expressions can serve as useful models for gas surface reactions, but should not be interpreted as proof of a particular reaction mechanism. Indeed, CVD analyses are complicated by gas phase reactions that form precursors for surface reactions; kinetic parameters are therefore seldom estimated.

TRANSPORT PHENOMENA - Similar to other chemically reacting systems such as combustion and heterogeneous catalysis, momentum, energy, and mass transport influence many aspects of CVD processes. For instance, the transport of reactants and impurities to and products from the growing film surface establishes film growth rates, composition, and uniformity. Temperature gradients and heat transfer also affect gas phase nucleation, film growth rates, and uniformity.

Transport processes occurring in CVD can be characterized by dimensionless parameter groups that result from scaling the governing transport equations [4,11]. For example, under most CVD reaction conditions, the gas velocities are low enough so that the fluid is in laminar flow as defined by a Reynolds Number, Re, where $\bar{V}$ is the average linear velocity,

$$Re = \frac{\bar{V}L}{\nu} \quad (4)$$

L is a characteristic distance (related to the reactor geometry) and ν is the kinematic viscosity. For typical CVD processes, Re ranges from 2-50.

Early flow visualization experiments led to the formulation of a simple boundary or stagnant layer model for transport effects in CVD [2]. However, recent experimental [13] and 3-D modeling [14] efforts have demonstrated that boundary layer flow does not occur in most CVD reactors. Depending upon the reactor configuration and deposition conditions, flow fields can be quite complex. For instance, large thermal gradients (arising from e.g., atmospheric pressure cold wall reactors with heated substrate holders) create buoyancy-driven free convection flows on the primary gas flow in the reactor. Since atmospheric pressure reactors generally operate in the mass transport-limited regime, achievement of uniform deposition rates requires that all substrates receive an equal reactant flux. This constraint limits throughput and makes process control difficult.

In order to minimize some of the adverse flow effects associated with atmospheric pressure CVD reactors, processes are often carried out at reduced pressure (~ 100 Pa) in hot wall (uniform system temperature) systems termed low pressure CVD (LPCVD) reactors. At constant mass flow, a pressure decrease increases the linear velocity and the mean free path of the gas molecules, leaves Re nearly constant, and substantially decreases buoyancy effects. Although these trends diminish the important of flow field considerations, the large mass diffusivities and essentially isothermal conditions cause diffusion and chemical reaction to be critical factors in process control [10]

Heat transfer is a crucial issue in many CVD processes. For instance, since LPCVD reactors are operated in the reaction rate controlled regime, even a few degrees variation in surface temperature can produce significant changes in deposition rate. Furthermore, the gas temperature must be accurately controlled in atmospheric pressure reactors to avoid homogeneous nucleation and thus particle generation. Finally, although conduction and convection are important heat transfer mechanics, the contribution from radiation is significant at the high temperatures (> 700 K) used in most CVD processes.

Mass transfer in CVD reactors occurs primarily through convection and diffusion due to concentration and temperature gradients. However, forced diffusion (e.g., electric field-driven ion transport) can be important in plasma-assisted CVD. Analogous to the Reynolds Number for momentum transport, the Peclet Number, Pe, gives an indication of the relative importance of convection versus concentration gradient diffusion,

$$Pe = \frac{\bar{V}L}{D} \quad (5)$$

where D is the diffusion coefficient. When $Pe \gg 1$, convection dominates; when $Pe \ll 1$, diffusion is the controlling process. For LPCVD processes, nearly equal contributions (i.e., $Pe \sim 1$) from convection and diffusion are observed.

Analogous to connective versus diffusive transport control, a CVD reaction may be dominated by diffusion or reaction kinetics.

Such comparisons are often made (in the chemical reaction engineering literature) by defining a Damköhler Number which indicates the relative importance of reaction to diffusion. Since D varies inversely with pressure, the controlling process in CVD may be switched from diffusion-limited to chemical reaction-limited as pressure is decreased. The deposition rate uniformity achieved under these conditions is the primary reason that LPCVD reactors are used.

CVD REACTOR MODELING

The goal of CVD reactor modeling is to relate process performance measures (e.g., deposition rate, film composition, uniformity) to operating conditions (e.g., temperature, reactant concentrations, reactor geometry, flow patterns) through a self consistent mathematical description of the underlying physicochemical phenomena. The formulation of models results in numerous practical advantages for CVD process development such as optimization, parameter estimation, equipment design, and predictive capabilities. In addition, however, modeling efforts should provide relationships between fundamental studies and process applications and generate new insight into the underlying chemical and physical principles involved.

A complete CVD reactor model combines thermochemical data, rate expressions, mass and energy fluxes, and continuity balances for momentum, chemical species, mass, and energy. The construction of such models is obviously a complicated exercise that requires sophisticated numerical methods to solve the appropriate coupled differential equations. Considerable progress has been made in the formulation of such models; however, these models are beyond the scope of this review. Interested readers can consult recent reviews of CVD modeling results [4,15].

CVD REACTOR TYPES

Due to the large parameter space in CVD processes, CVD reactors can be classified in various ways [1,2]. Typically, reactors are designated atmospheric pressure (APCVD), wherein a low partial pressure of reactants (i.e., a large carrier gas flow rate exists) is generally used, or low pressure (~ 100 Pa) in which high reactant partial pressures are invoked. Reactor temperatures are usually specified as high when they are above 773 K, and low below this somewhat arbitrary point. Furthermore, the wall temperature is labeled hot or cold depending upon whether the subtrate is at approximately the same or substantially below the temperature of the reactor walls. Finally, the flow configuration and substrate position are identified.

Schematics of the most commonly used CVD reactor configurations for microelectronic and optoelectronic device fabrication are shown in Fig. 3 [16]. Horizontal, vertical, barrel, and pancake (Figs. 1a-1d) CVD reactors are generally run under cold wall and near atmospheric pressure (LPCVD) conditions. Although the cooled walls minimize deposition on the walls and thus particulate and impurity problems, the large temperature gradients create secondary flow effects as described previously. In horizontal and barrel reactors, the substrate holders are tilted relative to the primary flow direction to improve film

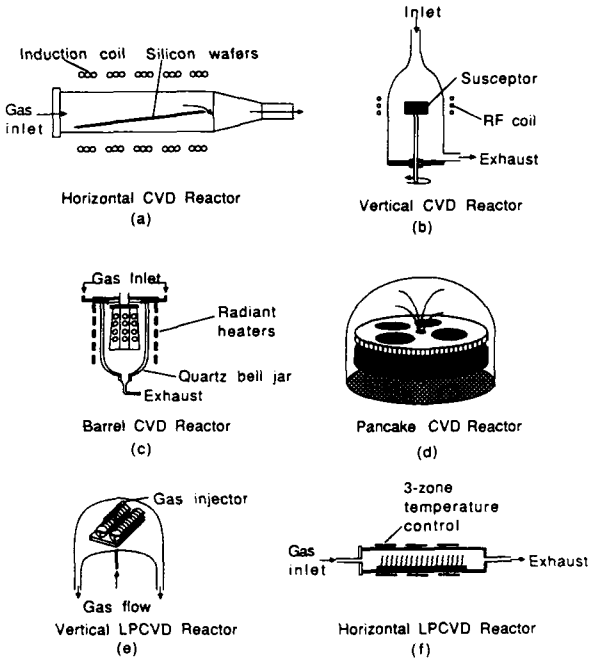


Fig. 3. Common CVD reactos [16].

uniformity. Additional uniformity control can be achieved by spinning the susceptor (1b), barrel (1c) or platen (1d) at speeds of 5-50 rpm. Unfortunately, the atmospheric pressure, cold wall reactors suffer from gas phase nucleation, which results in particulate contamination on substrates, and from generally poor step coverage over device topology.

LPCVD (30-250 Pa) hot wall reactors operating at temperatures between 820 and 1000 K can often overcome the above limitations of PACVD systems. Since mass transfer and fluid flow constraints are greatly reduced, close substrate spacing and thus high throughput is possible with LPCVD. Both vertical (to maximize uniformity) and horizontal flow configurations (Figs. 1e and 1f) have been used. Because reactor walls are

at essentially the same temperature as the substrates, film deposition occurs on the walls, thereby enhancing potential particulate problems. The low pressures utilized cause decreased deposition rates, although since diluents are generally not used, rates are only reduced by approximately a factor of 10.

EXAMPLES OF LPCVD PROCESSES

A broad spectrum of dielectric, semiconductor, and metallic thin films have been deposited by CVD [1]. The following sections briefly describe selected CVD processes utilized to form dielectric and metallic films employed extensively in microelectronic and optoelectronic applications. Since LPCVD is currently the most widely used approach to the deposition of these materials, emphasis will be placed on this method.

SILICON DIOXIDE - Deposited silicon dioxide (SiO_2) films are typically used as insulators between conducting layers and as ion implantation, diffusion, and outdiffusion masks. These films can be formed from a variety of reactant systems under different reactor conditions [2,17]. Structurally, the films are amorphous and at deposition temperatures below 773 K contain small amounts (< 2 atomic percent) of hydrogen bonded primarily as SiOH and SiH moieties. At low (< 700 K) temperatures, silane and oxygen can be reacted to deposit SiO_2 at atmospheric pressure or below.

$$SiH_4 + O_2 \rightarrow SiO_2 + 2\,H_2 \qquad (6)$$

Phosphorus-doped SiO_2 films, used as passivation and dielectric layers between metal films, can be formed by adding phosphine to the deposition atmosphere so that phosphorus oxide results.

$$4\,PH_3 + 5\,O_2 \rightarrow 2\,P_2O_5 + 6\,H_2 \qquad (7)$$

Poor step coverage and particle generation are the limitations of these low temperature processes.

At somewhat higher temperatures (920-1050 K) tetraethoxy silane vapor, $Si(OC_2H_5)_4$, termed TEOS, can be decomposed or reacted with oxygen to form SiO_2 and various organic and organosilicon byproducts [18,19]. Phosphorus and/or boron doping of these films can be performed by the addition of organic compounds containing dopant species to the deposition atmosphere [20]. The elevated temperatures and relatively stable reactant species means that the deposition is in the surface reaction rate limited regime, so that good step coverage and film uniformity are achieved. Unfortunately, in addition to the higher temperature requirement, the reaction chemistry is fairly complex and TEOS is a liquid which necessitates source and delivery line heating to avoid vapor condensation.

High temperature SiO_2 deposition can be carried out at ~ 1175 K and reduced pressure (~ 80 Pa) by the reaction between dichlorosilane and nitrous oxide [21,22].

$$SiCl_2H_2 + N_2O \rightarrow 2N_2 + 2HCl \qquad (8)$$

Although the dielectric quality of these films is quite good, and the deposition uniformity is excellent, chlorine incorporated into the deposited film can attack underlying silicon and can cause film cracking [22].

SILICON NITRIDE - Due to its high density, CVD silicon nitride is an excellent dielectric barrier to alkali ion and water vapor diffusion. Thus it is used as a high temperature protection or passivation layer against sodium ions and thermal oxidation. The films are amorphous and contain less than 8 atomic percent hydrogen, bonded to both silicon and nitrogen [23,24].

Silicon nitride can be deposited at atmospheric pressure by reacting silane and ammonia at temperatures between 973 and 1173 K. The ideal reaction may be represented by

$$3SiH_4 + 4NH_3 \rightarrow Si_3N_4 + 12H_2 \qquad (9)$$

Similarly, LPCVD with dichlorosilane and ammonia can be invoked at temperatures between 973 and 1073 K. The low pressure reaction is

$$3SiCl_2H_2 + 4NH_3 \rightarrow Si_3N_4 + 6HCl + 6H_2 \qquad (10)$$

favored due to excellent uniformity and high throughput. However, the reaction between dichlorisilane and ammonia generates ammonium chloride which can cause particulate problems. Furthermore, the high temperatures needed to obtain even low (5 nm/min) deposition rates are incompatible with low melting point films or substrates (e.g., aluminum, polymers).

TUNGSTEN - Tungsten films are of interest for metallic interconnection and contact/barrier layers in electronic devices to replace high resistivity polycrystalline silicon films and low temperature limitations and limited electromigration resistance aluminum [25-27]. However, the most advantageous property of LPCVD tungsten is the ability to selectively deposit this material on silicon (or metal) surfaces while surrounding silicon dioxide or silicon nitride regions do not nucleate tungsten film formation [25-27].

At temperatures below 570 K, tungsten can be selectively deposited by a two step process [28,29]. First, tungsten hexafluoride is reduced by a silicon surface.

$$2WF_6 + 3Si(s) \rightarrow 2W(s) + 3SiF_4 \qquad (11)$$

Tungsten layer growth is self-limiting, apparently due to diffusion of silicon through the W. The second step involves the hydrogen reduction of WF_6, usually at H_2/WF_6 ratios greater than 10.

$$WF_6 + 3H_2 \rightarrow W(s) + 6HF \qquad (12)$$

Above 575 K, selectivity falls off rapidly; under these conditions, it is possible to deposit a uniform continuous film. However, even at temperatures below 570 K, selectivity is difficult to maintain. Some nucleation often occurs on oxide (or nitride) surfaces during the process sequence.

TUNGSTEN SILICIDE - Lowered contact resistance and decreased resistivity compared to heavily doped polycrystalline silicon has been the driving force behind implementation of tungsten silicide in microelectronics processing [30]. Although the resistivity is not as low as tungsten, processing of tungsten silicide (WSi_2) is more compatible with current process technology since WSi_2 oxidizes in dry oxygen or water vapor at elevated temperatures to form a surface silicon oxide.

Tungsten silicide deposition can be performed under similar conditions to those of tungsten by merely adding silane to the deposition atmosphere [31].

$$WF_6 + 2SiH_4 \rightarrow WSi_2(s) + 6HF + H_2 \qquad (13)$$

A variation in WF_6/SiH_4 ratio permits a change in film composition (W/Si) and thus resistivity [31]. Alternatively, higher deposition rates and lower film fluorine content can be obtained by using disilane (Si_2H_6) as the silicon source instead of SiH_4 [32]. At the same substrate temperature, more silicon is incorporated into deposited films with Si_2H_6 compared to SiH_4 [32].

ALUMINUM - Due to the low resistivity and excellent adhesion to dielectrics, aluminum films are widely used as metallization layers in microelectronic devices [27]. To improve step coverage, LPCVD is of interest for Al films.

A number of aluminum compounds (mostly organic) have been used to deposit Al films [27], but the primary effort has been devoted to tri-isobutyl aluminum ($i\text{-}(C_4H_9)_3Al$) because of the relative ease of dissociation [33]. The principle problems involve film nucleation, adhesion, and rough surfaces [27], which currently preclude implementation of this process in production.

PLASMA AND PHOTON ASSISTED CVD

In CVD, reaction rates can be altered by varying the deposition temperature. When low melting point metals (e.g., aluminum) or polymers are present on a substrate, temperatures below 600 K must be maintained. This restriction can be met if the energy needed for bond breaking is supplied by energetic radiation such as electrons or photons.

PLASMA-ASSISTED CVD - Plasma-assisted CVD (PACVD) uses the high energy electrons present in glow discharges to dissociate and ionize gaseous molecules, thereby forming chemically reactive neutral species and ions for film deposition [4,34-37]. Since thermal energy is not needed to break chemical bonds, reactions can be promoted at room temperature. In addition to the numerous operating parameters that characterize CVD processes, PACVD has in addition, a need to control rf power and frequency. This means that the chemistry and physics of a glow discharge are extraordinarily complex. Nevertheless, the plasma performs only two basic functions. First, reactive chemical species are generated by electron impact collisions, thereby overcoming kinetic limitations that often exist in thermal CVD processes. Second, the discharge supplies energetic radiation (i.e., electrons, photons, and ions) that alters surface reaction steps (adsorption, reaction, and desorption). Thus, in addition to deposition rate changes, radiation bombardment of growing surfaces alters film composition, bonding structure, and ultimately properties. PACVD has been performed in essentially four different reactor configurations as shown in Fig. 4 [38]. In "capacitively coupled" systems (Figs. 4a and 4b), two parallel plate electrodes are positioned within a vacuum chamber. Substrates are generally placed on the grounded electrode which can positioned horizontally (Fig. 4a) or vertically (Fig. 4b). In both cases, the substrates are immersed in the discharge, and are thus subjected to ion, electron, and photon bombardment. "Inductively coupled" reactors (Figs. 4c and 4d) are often used when external electrodes are desired (for cleanliness purposes) or when higher temperatures ($\gtrsim$ 800 K) are needed. In these instances, the substrates may lie outside of (Fig. 4c) or within (Fig. 4d) the discharge volume. The downstream or "remote" configuration (Fig. 4c) is especially interesting because of the flexibility it presents to control or modify the specific chemistry that occurs [39,40]. For instance, all reactant gases can be passed through the discharge and allowed to react at the heated substrate surface. Alternatively, one (or more) of the reactants (or diluents) can be excited or dissociated by the plasma, and the reactive fragments brought into contact with other reactant molecules at the substrate surface. In this way, particular chemical bonds can be broken or molecules excited so that improved discrimination may result in the reaction chemistry and thus in the film bonding structure.

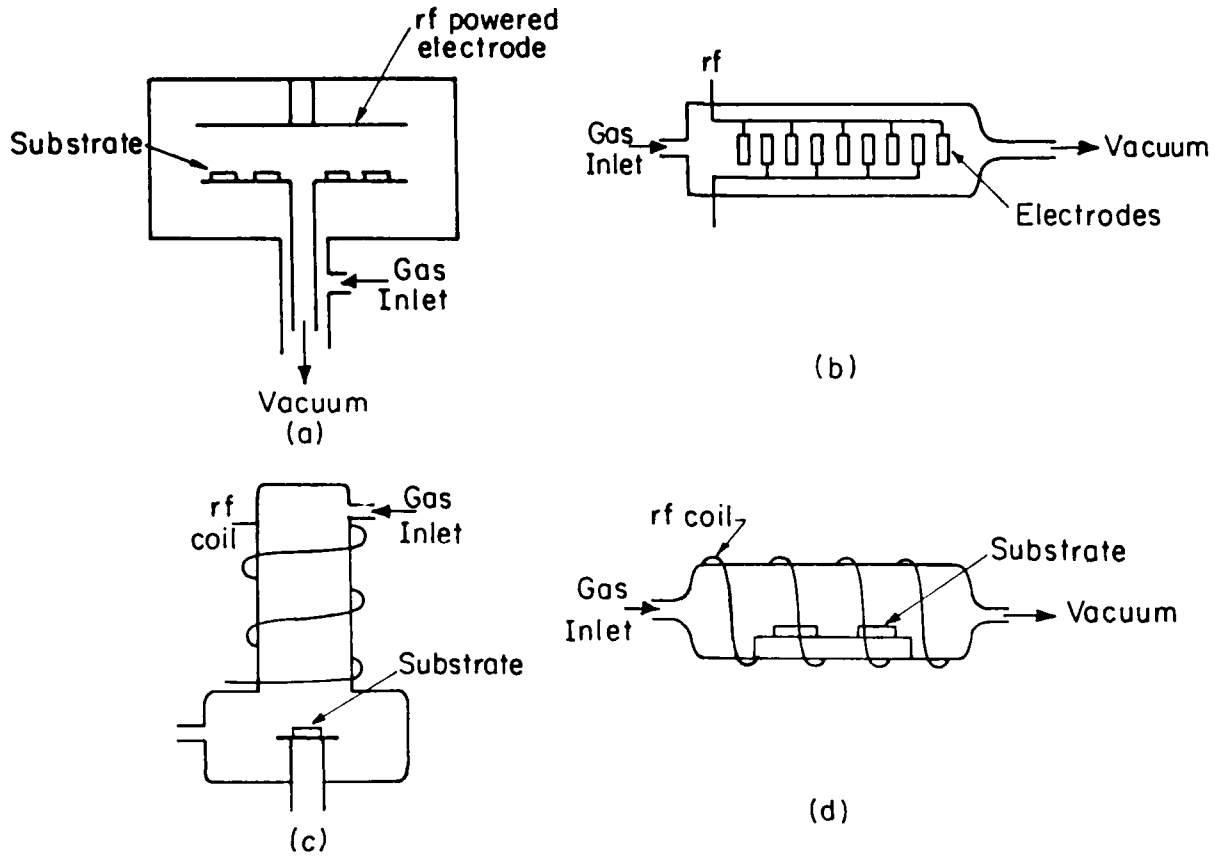


Fig. 4. Basic configurations of PACVD reactors [38].

Finally, various excitation sources (e.g., rf, microwave) can be used to optimize species concentrations so that deposition rates can be maximized.

The rf glow discharges invoked for PACVD are partially ionized gases composed of ions, electrons, and a host of neutral species in both ground and excited states. Application of an rf voltage at frequencies between 50 kHz and 40 MHz to a low pressure (6-600 Pa) gas results in a chemically unique environment. Electron, and because the plasma is electrically neutral, positive ion densities are in the range of 10^{8} - 10^{12} cm^{-3}. However, the ratio of neutral species density to the electron density is generally greater than 10^{3}, so that these are weakly ionized plasmas. As a result, neutral radicals and molecules in the discharge are primarily responsible for deposition reactions. That is, radicals and molecules are not inherently more reactive than ions, but are present in significantly higher concentrations. Because the average electron energy (> 2 eV) is considerably higher than ion (or thermal) energies (~ 0.04 eV), these discharges are termed non-equilibrium plasmas. The discharge therefore cannot be adequately described by a single temperature.

Due to the difference in mobility between electrons and positive ions, surfaces in contact with a discharge assume a negative potential with respect to the plasma [41]. This difference in potential results in an acceleration of positive ions into the growing film surface. Due to the momentum of positive ions or fast neutral species, these particles are quite effective in promoting surface reactions [42]. The ions are accelerated to energies between a few electron volts and several hundred electron volts, resulting in bond breaking and often in sputtering of surface atoms and molecules.

Deposited film properties are greatly altered by energetic particles, and to a lesser extent, by photon bombardment [43,44]. The plasma-surface phenomena that can result from particle impingement include surface cleaning, nucleation site generation, surface mobility enhancement, ion mixing, impurity incorporation, and defect generation. Clearly, these phenomena affect film growth

rates, impurity diffusion, and chemical, mechanical, and electrical properties.

As in CVD, a mathematical model for PACVD processes would be useful to assist the systematic and comprehensive understanding of these complex reactive atmospheres. In addition to the relatively unknown gas phase and surface chemistry in discharges, a number of complications exist compared to LPCVD processes. These problems include non-Maxwell-Boltzmann electron energy distributions, charged species transport and reactions, and alteration of surface chemistry by particle and photon bombardment. Because of such difficulties, PACVD modeling is less developed than that for conventional CVD. Nevertheless, some modeling efforts do exist, generally using approaches such as solving the Boltzmann equation, Monte Carlo methods, and "fluid" models. Although these techniques are beyond the scope of this review, information can be found in other sources [45-48].

PHOTON-ASSISTED CVD - Ultraviolet, visible, and infrared photons can be used in various ways to promote CVD reactions [4,49-54]. Typically, reactions are carried out at pressures between 15 and 200 Pa, and substrate temperatures between 320 and 600 K. Photon or laser irradiation of gas phase molecules, or the surface can promote deposition reactions. In the gas phase, photons can cause electronic excitation, thereby promoting reactions due to either an increased reactivity of excited state species, or the creation of reactive radicals by photo-dissociation. Indirect homogeneous excitation can be performed by sensitizing a vapor that does not participate in reactions except to transfer its excited state energy to reactants by collisional processes. Vibrational excitation of gas phase or adsorbed species can also enhance reactivity at a solid surface. Laser radiation absorbed by a solid surface can alter adsorption/desorption processes on the surface, or when relaxation occurs, can be transformed into heat. When heating is prevalent, essentially a thermal CVD process takes place, although only a localized surface region is affected.

EXAMPLES OF PLASMA-ASSISTED CVD PROCESSES

A large number of dielectric, semiconductor, and conducting thin films have been deposited by PACVD [34,36,55,56]. To date, little fundamental understanding of these complex processes is available; again, empirical studies have been the norm. The following two sections briefly describe the most extensively studied dielectric films deposited by PACVD.

SILICON NITRIDE - Since high temperature (> 1000 K) CVD silicon nitride is a dense, stable, adherent dielectric, PACVD silicon nitride, deposited at low (< 600 K) temperatures, is of interest as a protection coating and interlevel metal dielectric layer [57-60]. The deposition reaction is usually performed with SiH_4/NH_3 or SiH_4/N_2 reactants. Therefore, a critical aspect of PACVD silicon nitride is that a significant concentration of hydrogen (10-30 atomic %) is present in the deposited films. Most of the hydrogen is bonded to either silicon (~ 75%) or nitrogen (~ 25%). The specific concentration and chemical distribution of hydrogen greatly affects film properties such as refractive index, stress, density, and resistivity. Elevated (> 520 K) temperatures and high bombardment energies and fluxes (usually achieved at low frequencies) cause a reduction in hydrogen content and a decrease in the Si/N ratio, both of which establish film properties.

SILICON DIOXIDE - PACVD silicon dioxide films have been deposited by the reaction of SiH_4 with nitrous oxide (N_2O) or occasionally with carbon dioxide, carbon monoxide, or oxygen [61-64]. As in CVD, diborane or phosphine can be added to the deposition atmosphere to form doped oxide layers. Analogous to silicon nitride deposition, PACVD SiO_2 films contain hydrogen. Due to the enhanced reactivity of oxygen compared to nitrogen species with SiH_x fragments, lower concentrations of hydrogen are present in SiO_2 (2-9 atomic %)3 compared to silicon nitride.

The primary bonding configurations are SiH, SiOH, and H_2O. Naturally, the distribution of hydrogen between these moieties is dependent upon deposition conditions, primarily the deposition temperature. Furthermore, film properties depend on all plasma parameters. Finally, due to the use of N_2O in SiO_2 deposition, a small amount (< 5 atomic %) of nitrogen is incorporated in the deposited films.

SUMMARY AND CONCLUSIONS

Chemical vapor deposition (CVD) is a unique, flexible method for the deposition of an extensive variety of thin film materials from gas phase reactants. Even nonconventional CVD techniques (plasma - and photon-assisted) are widely used when low (< 600 K) temperatures are necessary. However, an understanding of the fundamental chemistry and chemical engineering principles is lacking. In particular, CVD kinetics require the same type of experimental and theoretical studies that have enhanced our knowledge of combustion chemistry and heterogeneous catalysis. Rate coefficients and accurate kinetic expressions for gas phase, electron impact, and surface reactions are generally unknown for the chemical reactions used in the formation of thin films. Transient (e.g., fluid flow, mass transport) phenomena occurring at the start or the end of film deposition cycles are not understood. A basic understanding of

nucleation phenomena is also missing. Generation of these data and information will permit the formulation of precise kinetic models for deposition processes. Ultimately, the establishment of such models will lead to improved reactor design concepts and scale-up procedures, thereby eliminating the current trail-and-error approaches. Only when the above problems are solved, will CVD attain its full potential as a controllable, reproducible film formation method.

ACKNOWLEDGEMENT

The author appreciates the helpful comments and critical reading of the manuscript by Professor David Graves.

REFERENCES

[1]. W. Kern and V. Ban, in "Thin Film Processes" (J. L. Vossen and W. Kern eds.) pp. 257-331, Academic Press, New York (1978).

[2]. A. C. Adams, in "VLSI Technology" (S. M Sze ed.) pp. 93-129, McGraw-Hill, New York (1983).

[3]. McD. Robinson, C.H.J. van den Brekel, G. W. Cullen, J. M. Blocher, Jr., and P. Rai-Choudhury, eds. "Proc. Ninth Int'l Conf. on CVD," The Electrochem. Soc., Pennington (1984).

[4]. Hess, D. W., K. F. Jensen, and T. J. Anderson, Reviews in Chem. Eng., 3, 97-186 (1985).

[5]. G. W. Cullen and J. M. Blocher, Jr., eds., "Proc. Tenth Int'l Conf. on CVD," The Electrochem. Soc., Pennington (1987).

[6]. Jasinski, J. M., B. S. Meyerson, and B. A. Scott., Ann. Rev. Phys. Chem. 38, 109-140 (1987).

[7]. F. van Zeggeren and S. H. Storey, "The Computation of Chemical Equilibria," Cambridge University Press, Cambridge (1970).

[8]. W. R. Smith and R. M. Missen, "Chemical Reaction Analysis," Wiley Interscience, New York (1982).

[9]. J. J. Carberry, "Chemical and Catalytic Reaction Engineering," Chap. 8, McGraw-Hill, New York (1976).

[10]. Roenigk, K. F., and K. F. Jensen, J. Electrochem. Soc., 134, 1777-1785 (1987).

[11]. F. Rosenberger, in "Proc. Tenth Int'l Conf. on CVD," (G. W. Cullen and J. M. Blocher Jr., eds.) pp. 11-22, The Electrochem. Soc., Pennington (1987).

[12]. Eversteyn, F. C., P.J.W. Severin, C.H.J. van den Brekel, and H. L. Peek, J. Electrochem. Soc., 117, 925-931 (1980).

[13]. Giling, L. G., J. Electrochem. 129, 634-644 (1982).

[14]. Moffat, H. K., and K. F. Jensen, J. Electrochem. Soc., 135, 459-471 (1988).

[15]. Jensen, K. F. Chem. Eng. Sci. 42, 923-958 (1987).

[16]. Jensen, K. F. to be published in "Chemical Engineering in Electronic Materials Processing" (D. W. Hess and K. F. Jensen, eds), American Chemical Society Symposium Series No. 221.

[17]. Kern, W., and R. S. Rosler, J. Vac. Sci. Technol., 14, 1082-1099 (1977).

[18]. Adams, A. C., and C. D. Capio, J. Electrochem. Soc., 126, 1042-1046 (1979).

[19]. Levin, R. M., and A. C. Adams, J. Electrochem. Soc., 129, 1588-1592 (1982).

[20]. Williams, D. S., and E. A. Dein, J. Electrochem. Soc., 134, 657-664 (1987).

[21]. Rosler, R. S., Solid State Technol., 20(4), 63-70 (1977).

[22]. Watanabe, K., T. Tanigaki, and S. Wakayama, J. Electrochem. Soc., 128, 2630-2635 (1981).

[23]. Makino, T., J. Electrochem. Soc., 130, 450-455 (1983).

[24]. Pan, P., and W. Berry, J. Electrochem. Soc., 132, 3001-3005 (1985).

[25]. Blewer, R. S., Solid State Technol. 29(11), 117-126 (1986).

[26]. Itoh, H., T. Moriya, and M. Kashiwagi, Solid State Technol., 30(11), 83-87 (1987).

[27]. Levy, R. A., and M. L. Green, J. Electrochem. Soc., 134, 37C-49C (1987).

[28]. Pauleau, Y., and Ph. Lami, J. Electrochem. Soc., 132, 2779-2784 (1985).

[29]. McConica, C. M., and K. Krishnamani, J. Electrochem. Soc., 133, 2542-2548 (1986).

[30]. S. P. Murarka, "Silicides for VLSI Applications," Academic Press, Orlando (1983).

[31] Brors, D. L., J. A. Fair, K. A. Monnig, and K. C. Saraswat, Solid State Technol. 26(4), 183-186 (1983).

[32]. Shioya, Y., K. Ikegami, I. Kobayashi, and M. Maeda, J. Electrochem. Soc., 134, 1220-1224 (1987)

[33]. Cooke, M. J., R. A. Heinecke, R. C. Stern, and J. W. Maes, Solid State Technol., 25(12), 62-65 (1982).

[34]. J. R., Hollahan and R. S. Rosler, in "Thin Film Processes" (J. L. Vossen and W. Kern, eds.), pp. 335-360, Academic Press, New York (1978).

[35]. Thorton, J. A., Thin Solid Films, 107, 3-19 (1983).

[36]. Hess, D. W., J. Vac. Sci. Technol., A2, 244-252 (1984).

[37]. Y. Catherine, in "Plasma Procesing" (G. S. Mathad, G. C. Schwartz, and G. Smolinsky eds.), pp. 317-344, Electrochem. Soc., Inc., Pennington (1985).

[38]. D. W. Hess, in "Silicon Processing" (D. C. Gupta, ed.), pp. 218-225, Amer. Soc. Testing and Mater., Washington, D.C. (1983).

[39]. Meiners, L. G., J. Vac. Sci. Technol., 21, 655-658 (1982).

[40]. G. Lucovsky, D. V. Tsu, and R. J. Markunas, in "Plasma Processing" (J. W. Coburn, R. A. Gottscho, and D. W. Hess, eds.), pp. 323-334, Materials Res. Soc., Inc., Pittsburgh (1986).

[41]. Vossen, J. L. J. Electrochem., Soc., 126, 319-324 (1979).

[42]. Gerlach-Meyer, U. J. W. Coburn, and E. Kay, Surf. Sci., 103, 177-188 (1981).

[43]. Takagi, T., J. Vac. Sci. Technol., A2, 382-388 (1984).

[44]. Hess, D. W., Ann. Rev. Mat. Sci., 16, 163-183 (1986).

[45]. Turban, G., Y. Catherine, and B. Grolleau, Plasma Chem. Plasma Proc., 2, 61-80 (1982).

[46]. Kushner, M. J., IEEE Trans. Plasma Sci., PS-14, 188-196 (1986).

[47]. Rhee, S., and J. Szekely, J. Electrochem. Soc., 133, 2194-2201 (1986).

[48]. D. B. Graves, in "Plasma Processing" (G. S. Mathad, G. C. Schwartz, and R. A. Gottscho, eds.) pp. 267-288, Electrochem. Soc. Inc., Pennington (1987).

[49]. Chuang, T. J., J. Vac. Sci. Technol., 21, 798-806 (1982).

[50]. S. D. Allen, in "Physical Processes in Laser-Materials Interactions" (M. Bertolotti, ed.), pp. 455-468, Plenum Press, New York (1983).

[51]. Ehrich, D. J., and J. Y. Tsao, J. Vac. Technol., B1, 969-984 (1983).

[52]. Osgood, R. M., Jr., Ann. Rev. Phys. Chem., 34, 77-101 (1983).

[53]. P. K. Boyer, C. A. Moore, R. Solanki, W. K. Ritchie, G. A. Roche, and G. J. Collins, in "Laser Diagnostics and Photochemical Processing for Semiconductor Devices" (R. M. Osgood Jr., and S. R. J. Brueck, eds), pp. 119-127, North Holland, Amsterdam (1983).

[54]. Hargis, P. J., and J. M. Gee, Solid State Technol., 21(11), 127-133 (1984).

[55]. D. W. Hess, in "Reduced Temperature Processing for VLSI" (R. Reif and G. R. Srinivasan, eds.), pp. 3-23, Electrochem. Soc., Inc. Pennington (1986).

[56]. A. C. Adams, in "Reduced Temperature Processing for VLSI" (R. Reif and G. R. Srinivasan, eds.), pp. 111-131, Electrochem. Soc., Inc., Pennington (1986).

[57]. Claassen, W.A.P., W.G.J.N. Valkenburg, M.F.C. Willemsen, and W.M. v.d. Wijert, J. Electrochem. Soc., 132, 893-898 (1985).

[58]. Maeda, M., and H. Nakamura, J. Appl. Phys., 58, 484-489 (1985).

[59]. Knolle, W. R., and J. W. Osenbach, J. Appl. Phys., 58, 1248-1254 (1985).

[60]. Livengood, R. E., and D. W. Hess, Appl. Phys. Lett., 50, 560-562 (1987).

[61]. Reinberg, A. R., J. Electron. Mat., 8, 345-372 (1979).

[62]. Adams, A. C., Solid State Technol., 26(4), 135-139 (1983).

[63]. T. B. Gorczyca and B. Gorowitz, in "VLSI Electronics: Microstructure Science", Vol 8 (N. G. Einspruch and D. M. Brown, eds.), pp. 69-88, Academic Press, New York (1984).

[64]. Nguyen, S. V., J. Vac. Sci. Technol., B4, 1159-1167 (1986).

TRANSITION METAL SILICIDES FOR MICROELECTRONICS

Sung-Joon Kim
California Institute of Technology
Pasadena, California 91125, USA
Present Address: AT&T Bell Laboratories
Murray Hill, New Jersey 07974, USA

Marc-A. Nicolet
California Institute of Technology
Pasadena, California 91125, USA

ABSTRACT

Transition metal silicides have long been scientifically interesting and technologically desirable because of their properties for metallurgical and electronic applications. Especially in the last two decades, silicides have been investigated extensively in the field of microelectronics. What makes them attractive is their metal-like conductivity, high temperature stability, oxidizability, and compatibility with integrated circuit processing conditions. Silicides have been used so far in integrated circuits for interconnections, Schottky barriers and ohmic contacts. With the recent advancements in thin film science, advanced usages of silicides are now attempted that exploit the existence of pseudomorphic phases of certain silicides on single crystal silicon.

We briefly review here applications, formation, and properties of silicides that are important in integrated circuit technology with the emphasis on recent progress.

SILICIDES HAVE LONG BEEN INVESTIGATED for various electronic applications. Table I summarizes some of their uses. With the advance of integrated circuit (IC) technology in the last two decades, the role of silicides in microelectronics has become prominent. Research leading to a practical usage of silicides as Schottky barriers, as ohmic contacts, and as interconnection lines in IC has been very intense. Numerous review articles have been published on this subject in the past (for listings, see i.e. refs. 1 and 2). These papers provide good summaries of the properties of silicides that are relevant to electronic applications. In the present article, we concentrate on the way silicides are used in microelectronics and emphasize some of the most recent developments.

II. Refractory Metal Silicides for Gate Materials in MOS Devices

In the late 60's doped polycrystalline-silicon (poly-Si) was the common material for gate electrodes and interconnections in MOS devices. However, the downscaling of the dimensions of devices and circuits in the 70's to increase the packing density of IC made the lowering of the electrical resistance of the gate and interconnection material a rising and critical issue. What was needed was a material with an electrical resistivity that was significantly less than that of doped poly-Si (300 $\mu\Omega$ cm). Other requirements were a small grain size to ease high resolution patterning, high temperature stability, good passivation, resistance to chemicals used in the IC processings, and the formation of good ohmic contacts to other levels of metallizations.

Refractory metal silicides, satisfy most of these requirements. Since in the late 70's, the study on the applications of refractory metal silicides has been intensive and many review papers have been published [2,3]. Table I lists some actual and proposed applications of silicides that have come out of these efforts. We note in passing that silicides have long been known in metallurgy for their properties as protective coatings (refs. 18,19 in Table I).
The refractory metal silicides which are considered here are the dislicides of Mo, W, Ta, Ti and Nb. Their resistivities range from $15\sim(TiSi_2)$ to ~ 100 $(MoSi_2)$ $\mu\Omega$ cm. Resistivities and other properties of these silicides depend on the formation method. Films that are formed by thermal annealing from coevaporation generally have lower resistivities than the ones from co-sputtering; films formed by co-sputtering results in better quality than the ones by sputtering from composite targets; films formed by chemical vapor deposition (CVD) have better adhesion to SiO_2 substrate than the ones deposited by sputtering or evaporation. The choice of the formation method depends on

TABLE I. ACTUAL AND PROPOSED APPLICATIONS OF SILICIDES

*APPLICATIONS	SILICIDES	RELEVANT AND REQUIRED PROPERTIES	*METHODS OF FORMATION	REFERENCES
Gate and interconnection lines in Si integrated circuit	W,Mo,Ta, and Ti disilicide	Low resistivity (20-100 μΩcm),high temperature stability, oxidizability	Cosputtering or CVD and anneal by CFA or RTA	1
Shallow junction contacts, gate and interconnection lines in Si integrated circuit	Ti,Co,Ni disilicides and PtSi, Pd_2Si	Lower resistivity than refractory metal silicides with the exception of $TiSi_2$.Compatible with self-aligned silicide (salicide) process, oxidizability, reliable contacts and Schottky barrier.	Metal reaction with Si by CFA or RTA	1
Metal and permeable base transistor	$CoSi_2$,$NiSi_2$	Low resistivity, epitaxial silicide growth on Si and epitaxial Si growth on silicide	MBE,SPE, high dose metal implantation	2,3,4,5
Integrated semiconductor negative resistance diode; integrated multiple IMPATTs	$CoSi_2$	same as above	MBE, SPE	6
IR photodectors	$CoSi_2$,$NiSi_2$, Pd_2Si,PtSi, $CrSi_2$, $MnSi_{1.7}$, $IrSi_{1.75}$,	Metallic or semiconducting silicides and epitaxial silicide growth on Si.	MBE, SPE	6,7,8,9, 10
Light source and detectors	$FeSi_2$	Semiconducting silicides with direct bandgap. Epitaxy growth of silicides on Si.	MBE	9
Thermomodules	$FeSi_2$,$CrSi_2$, CoSi,	Semiconducting and semimetal silicides	Solid state reaction	12,13
Self-aligned gate on GaAs	WSi_2(W_xSi_3) ($0.6<x<0.8$) TiWSi ($x\geq0.4$) $TaSi_x$	High temperature stability optimal device electrical properties planar process, reduction of excessive parasitic source resistance.	Sputtering	14,15,16
Archival optical storage	PdSi, PtSi RhSi	Optical reflectivity change before and after silicide formation, long lifetime of reflectivity, nontoxic large contrast between silicide and the metal. Low-power laser interaction	Laser heating	17
Protection coating for refractory metals from high temperature oxidation	$MoSi_2$, WSi_2, Ta_5Si_3, Nb_5Si_3	High temperature stability, oxidation resistant	Gas metal reaction using siline	18,19,

*CVD - Chemical vapor deposition
MBE - Molecular beam epitaxy
SPE - Solid phase epitaxy
IMPATT - Impact avalanche transit time device
RTA - Rapid thermal annealing
IR - Infrared
FET - Field effect transistor
MESFET - Metal semiconductor field effect transistor
CFA - Conventional furnace annealing
MODFET - Modulation doped field effect transistor

Table II. TYPES OF SILICIDE GATES FOR MOS DEVICES

TYPES	ADVANTAGES	DISADVANTAGES	SILICIDES	REFERENCES
Single-level silicide/SiO_2	Low-resistivity when silicide is not consumed during oxidation. Si-rich silicide can alleviate this problem.	Interfacial instability with underlying SiO_2. Silicide deteriorates after oxidation.	$MoSi_2$, WSi_2, $NbSi_2$	1,2
Deposited silicide/poly gate Si (polycide)/SiO_2	Gate integrity preserved after oxidation. Good poly-Si/SiO_2 interface stability. Low-resistivity and high reliability	Added processing steps, i.e. Silicide is deposited on top of poly-Si. High resolution patterning difficult. Possible lateral undercutting	WSi_2, $MoSi_2$, $TaSi_2$, $TiSi_2$, $NbSi_2$	1,2
Silicided of metal/metal/SiO_2	Low resistivity metal combined with silicide for desirable chemical reactions (i.e. metal oxide formation is prevented)	Instability between metal and its silicides.	WSi_2, $MoSi_2$	1,2
Silicided poly-Si/poly-Si/SiO_2 (salicide)	Source/drain can also be silicidized at the same time. Self-aligned structure. Easy oxidation.	Dopant redistribution during silicide formation.	PtSi, $TiSi_2$, $MoSi_2$, WSi_2	1,2
Metal/Silicide/poly-Si/SiO_2	High temperature stability between metal and poly-Si. Low-resistivity metal interconnection structure	Increased number of critical processing steps.	Mo/$Mosi_x$	3

Table III. PROPERTIES OF REFRACTORY AND NEAR-NOBLE METAL SILICIDES

PROPERTIES	REFRACTORY METAL SILICIDES (WSi_2,$MoSi_2$,$TaSi_2$,$TiSi_2$,$NbSi_2$)	NEAR-NOBLE METAL DISILICIDES ($CoSi_2$,$NiSi_2$,PtSi,Pd_2Si)
Resistivities ($\mu\Omega$cm)	14 - 120	10 - 50
Schottky barrier on n-Si (eV)	0.59 - 0.65	0.64 - 0.88
Dominant diffusion species	Si	Metal
Activation energy of growth by reaction with metal film (eV)	2.5 - 3.7	0.9 - 1.5
Growth rate	Probably diffusion-limited, but severly affected by impurities, particularly the silicides of Ta. Mo and W which are quite difficult to form by metal-Si reaction; they are normally prepared by co-deposition and annealing.	PtSi,PdSi-Diffusion limited $NiSi_2$-Nucleation limited $CoSi_2$-Nucleation and diffusion limited
Temperatures at which silicide film on Si substrate is unstable (°C)	950 - 1000	700 - 900

considerations of cost and the particular optimum of properties desired.

There are various ways silicide films can be used in a gate of a MOS device. They are summarized in Table II and also shown in Fig. 1. Each type of gate has its own advantages and drawbacks. For instance, a silicide/poly-Si gate has good device properties and structural integrity, but, patterning it uniformly by conventional wet etch is hard since the etching rates of poly-Si and silicide are different. On the other hand, a single-level silicide can be patterned easily but the film is not stable with the underlying SiO_2 layer. The polycide configuration is most commonly used in micron-size devices since this configuration incorporates the advantages of both silicide and poly-Si properties as described in Table II. As the processing technologies continue to improve and the new ones are developed, a new type of gate design would likely be used for submicron devices. These devices require an accuracy in the definition in the pattern that cannot be attained at present with polycides.

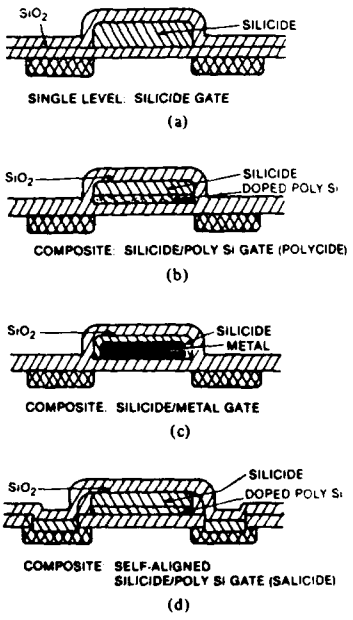


Fig. 1. Schematic cross sections of various refractory silicide gate structures: (a) single-level silicide, (b) "polycide," (c) silicidation of metal and (d) silicidation of poly-Si ("salicide") [3].

III. Silicides for Shallow Junction

The shrinking of the size of the device to increase the density of devices on a chip has also resulted in the decrease of junction depth to the range of a few tens of nanometers. Aluminum which was used as a contact material before the submicron era, is no longer directly applicable with shallow junctions. After Al is deposited on Si, there follows an annealing step to sinter the contact and stabilize it. During this process Al and Si interdiffuse. For shallow junction contacts, this process is no longer acceptable. Silicides which have a good stability in contact with Si under relatively high temperature annealing, are an obvious choice to replacing Al for forming contacts to shallow junctions.

When a silicide is formed by reaction of a metal film with a Si substrate, a new interface is created between the Si substrate and the penetrating silicide layer. Impurities, crystalline defects and native SiO_2 at the original Al/Si interface are consumed in the process by the growing silicide layer. As a consequence, the interface between silicides and Si yields Schottky diodes or ohmic contacts that are of superior reliablity and reproducibility.

However, the most desirable way of forming a chemically and electronically reliable silicide-Si contact would be by forming an epitaxial silicide on defect and oxide free Si substrate by processing it in an ultra-high vacuum system. Formation and characteristics of epitaxial silicides will be discussed later in section V.

Whether a silicide-Si contact is rectifying or ohmic depends on the doping concentration of Si substrate. On lightly doped n-type Si substrates, a silicide contact is typically rectifying. With high doping concentration of either type (above $10^{20}/cm^3$), the contacts are ohmic and they are used for the source and drain electrodes of MOS transistors.

The types of silicides used in these applications depends on the processing requirements (annealing temperatures, passivation conditions, etching process, reactivity with surrounding environment), and on the properties of the silicide (electrical resistivity, barrier height of silicides to Si). Some of the characteristics of two important groups of silicides for metallization application are listed in Table III [1,4,5]. Refractory metal silicides, except $TiSi_2$, are best formed by co-deposition and subsequent annealing. For near-noble metal silicides and $TiSi_2$, metal films can be deposited on the Si substrate and thermally reacted with the Si substrate. Near-noble metal silicides and $TiSi_2$ have lower resistivities than refractory metal silicides, but refractory metal silicides are more stable in contact with Si at high temperatures than near-noble metal silicides. From all that precedes, near-noble metal silicides are the best choice for Si contacts if the processing

temperature can be kept low or if the metallization can be applied after all high temperature cycles are completed.

IV. Self-Aligned Silicide (Salicide)

Among the four configurations of gate designs for MOSFETs shown in Table II, the self-aligned silicide (salicide) process is of particular interest especially in forming shallow junctions for submicron devices. In this process silicides are formed simultaneously at gate, source and drain contact levels. This is an advantage over the separate application of silicides on the gate and the source and drain contacts for MOS devices, since the number of processing steps is reduced. In Fig. 2 a typical sequence of salicide processing steps is shown [6]. Similar processing steps also apply for contact formation in bipolar devices [7].

The main steps in this process are; 1) deposit silicon on gate oxide and dope it heavily; 2) place an oxide at the sidewalls of the gate by depositing a blanket layer of SiO_2 by low pressure chemical vapor deposition (LPCVD) method and anisotropically etch it with a plasma to leave only the oxide sidewalls; 3) deposit a blanket metal film after the poly-Si and Si surfaces have been in-situ cleaned; 4) thermally anneal to form silicides. Duration, temperature, and ambient for this annealing step are critical because a lateral propagation of the silicide reaction could produce shorts between the gate and the source or drain contacts. Reaction products of the metal (i.e. Ti) with the SiO_2 sidewall can be difficult to etch; 5) etch the unreacted metal.

Some of the advantages of this process over the contact formed by co-deposition or multilayer deposition of metal and Si are; a) a uniform and clean silicide-Si interface is formed which
results in a reliable contact, b) the number of photolithographic steps are reduced since the silicide-forming process is self-aligned, c) an elemental metal film is deposited instead of a co-deposition from composite target, thereby reducing contaminations and composition ratio problems, and d) the complicated step of etching a silicide and poly-Si stack to form the gate in MOS devices is circumvented.

The salicide process is attractive, but it has problems of its own. During the silicide formation, and especially after subsequent high temperature processes, the dopants in the shallow Si junction region redistribute which can adversely affect the contact properties. The effect depends on the type of dopant and the silicide. For instance, in CMOS device processing with $TiSi_2$/Si contacts [6], a high temperature step after salicidation causes boron to segragate into the $TiSi_2$ and the contact resistance to rise. Arsenic in Si is not that highly diffusive at high temperature so that in n-type contacts, and in NMOS technology, shallow junctions remain unaffected. There is also a problem with the lateral spreading of the reaction that could electrically shorten the source or drain contacts with the gate. The duration, temperature, and ambient during silicide formation need to be carefully controlled to minimize this lateral spread. The vertical penetration of the silicide into the Si of the gate also has to be controlled. The direct contact of a silicide with the gate oxide of a MOS device generates stress which could result in the failured gate oxide integrity. Salicidized MOS devices have lower hot electron hardness than the non-silicidized devices [6]. Such problems are suspected to be due to current rounding near the shallow junction and the stress caused by the gate salicidation. The critical steps that also need special attention in salicide formation is the complete removal of unreacted metal from the oxide. Traces of unreacted metal on the oxide could result in bridging of contacts and gate.

Presently the most commonly used silicide for the salicide process is $TiSi_2$ [6,7,8]. $TiSi_2$ can be formed relatively easily by solid-state reaction of deposited Ti with Si, has low resistivity, and is stable at high temperatures. A drawback is that Ti reacts with SiO_2 near 600°C in addition to the silicide formation with Si.

Silicon is the dominant diffusing species during the formation of $TiSi_2$. This fact is at the root of several of the problems mentioned above and also leads to a loss of line width if the growth is not well controlled [9]. On the other hand, near-noble metal silicides form by the dominant diffusion of the metal and do not have the above problems but they are less stable at high temperatures than $TiSi_2$. The general trend, however, is to lower the temperatures in all VLSI processing to meet the requirements of

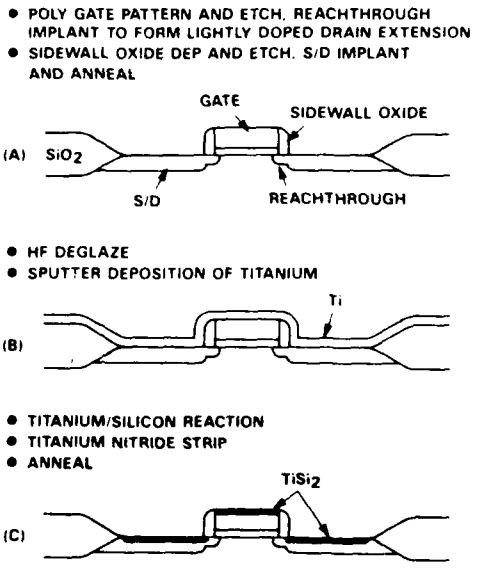


Fig. 2. Fabrication of MOS transistors with low resistivity gates and junctions using the self-aligned $TiSi_2$ process [6].

Table IV. SILICIDES AS SCHOTTKY CONTACTS TO GaAs-TYPE DEVICES

SILICIDE*	SUBSTRATE	METHOD OF PREPARATION	RESULTS	REFERENCES (same as text)
$TaSi_2$	$n(1.9\times10^{17}cm^{-3})/n^+$ GaAs	RF sputtering of compound target	Electrically stable after 800°C annealing. $n\sim1.1$ $\phi_B\sim0.79V$	15
$TaSi_2$	n-GaAs	same as above	After 650°C annealing As and/or Ga seem to migrate to $Tasi_2$ from RBS and Auger electron microscopy. No Ta or Si migration is detected after annealing for up to 850°C.	11
Ti/W Silicide	$n(2\times10^{17}cm^{-3})$GaAs	RF sputtering	From SIMS and RBS measurement, no evidence of reactions observed between Ti/W silicide and GaAs after 850°C annealing for 1 h. ϕ_B as-deposited ~0.75 V at 850°C ~0.82 V n as-deposited 1.05 n at 850°C ~1.2	10
$(TiW)Si_{0.4}$	n^+(Al,Ga)As/GaAs heterostructure	rf cosputtering of metals and Si from separate targets	After 800°C annealing no interdiffusion is observed from RBS and AES measurements	12
Ti-W-Si	$n(4\times10^{17}cm^{-3})$GaAs	Magnetron co-sputtering from separate targets	After 900°C annealing, I-V results are stable. n 1.15, $\phi_B\sim0.74$eV. After 850°C annealing, from C-V measurement; carrier concentration at 100nm depth has decreased to $1.6\times10^{17}cm^{-3}$ from before annealing value of $\sim4\times10^{17}cm^{-3}$. After 850°C annealing Ti diffuses into GaAs and Au diffuses out as from SIMS measurement.	13
W_3Si	$n(2\times10^{17}cm^{-3})$GaAs	same as above	After 850°C annealing C-V results are stable. Carrier concentration before and after annealing is $1.8\times10^{17}cm^{-3}$.	13
$W_{.9}Si_{.1}$	$n(2\times10^{17}cm^{-3})$GaAs	same as above	after 850°C annealing non-uniform C-V results. I-V results, however, are stable for before and after annealing. ($\Delta n\sim0.1, \Delta\phi_B\sim0.02$eV). SIMS measurement also show stable interface after 850°C annealing.	13
$W_{.55}Si_{.45}$	$n(2\times10^{17}cm^{-3})$GaAs	Magnetron cosputtering from separate targets	Very unstable C-V results after 850°C annealing. I-V characteristics are also unstable and the diode becomes leaky in reverse bias.	13
$WSi_{0.6}$ (W_5Si_3).	n^+(Al,Ga)As/GaAs. heterostructure		RBS and AES do not show any interfacial reaction after 800°C annealing.	12
$WSi_{0.6}$	$n[5.9\sim6.2\times10^{16}cm^{-3}$ GaAs	rf sputtering from composite target	as deposited: $n\sim1.03$ $\phi_B\sim0.64$eV n and ϕ_B degrade when annealed at 600-700°C	14

Table IV. SILICIDES AS SCHOTTKY CONTACTS TO GaAs-TYPE DEVICES (continued)

SILICIDE*	SUBSTRATE	METHOD OF PREPARATION	RESULTS	REFERENCES (same as text)
$WSi_{2.3}$		same as above	as deposited: $n \sim 1.07$, $\phi_B \sim 0.61$eV; n and ϕ_B degrade when annealed at 600-700°C	14
$WSi_{2.3}$		same as above	as deposited: $n \sim 1.04$, $\phi_B \sim 0.64$eV; n and ϕ_B degrade when annealed at 600-700°C.	14
$WSi_{0.6}$	n[5.9 $2x10^{16}cm^{-3}$] GaAs	rf sputtering from separate targets	As deposited and after 750°C annealing $n \sim 0.99$, $\phi_B \sim 0.79$eV	14

*n - I-V ideality factor
ϕ_B - Schottky barrier height

Table V. EPITAXIAL GROWTH OF RARE EARTH SILICIDES [$ReSi_{1.7}$] ON {111} Si AND EXPERIMENTAL ION CHANNELING MIMIMUM YIELDS

Formation methods	Y	Gd	Tb	Dy	Ho	Er	Tm	Yb	Lm	Ref.
					Xmin					
Metal evaporated on Si substrate held at 200°C. Epitaxial layer formed by solid state reaction from rapid e-beam heating.	26%		85%		81%	35%	60%	71%	78%	1,2
Metal evaporated on Si substrate held at 200°C. Epitaxial layer formed by solid state reaction from rapid halogen lamp	69%		90%	78%	44%	95%			72%	1,2,
Metal evaporated on Si substrate held at 200°C. Epitaxial layer formed by liquid phase reaction from rapid e-beam heating		93%		57%			75%	25%	20%	1,2,
Y is Ar sputtered on Si and conventional furnace annealing was conducted to form YSi_{2-x} ($0<X<0.3$). Use of Shiraki technique for Si substrate cleaning is critical for epitaxial growth	8%	($\rho \sim 50\mu\Omega$cm, $\phi_B \sim 0.36$ eV)								3

submicron technology for tight control of doping profiles over small dimensions. In that case, near noble metals with their low resistivities (see Table III), could become attractive alternatives for the salicide technology.

V. Gates to GaAs-type devices

GaAs MESFETs use Schottky contacts for the gate of the device. Both refractory and near-noble metals, such as Cr, Ti, W, Au and Al can be used as gate electrodes. These devices cannot go through high temperature processings such as annealings after ion-implantation (>800°C), because the Schottky contacts are observed to deteriorate. To implement the self-aligned source and drain technology by ion implantation after the gate electrode is deposited, gate materials are needed that can sustain high temperature processing. Refractory metal silicides such as WTi_ySi_x [10,12], TaSi [11,12], and WSi_x [12,13,14] have been investigated for that purpose as replacements for the elemental metal gates. These refractory metal silicides are sputtered from a single compound target, separate metal and Si targets or composite target (target A covered with strips of B). The Schottky contacts formed by these silicide layers to n-type GaAs have been characterized from the point of view of their mechanical, chemical and electrical stability upon thermal annealing. Table IV summarizes some of the results. The contacts have been characterized by Schottky barrier height, ϕ_B, and ideallity factor, n, from I-V measurement, depletion layer depth and carrier concentration from C-V measurement, and interfacial diffusion reaction from Rutherford backscattering spectrometry (RBS), secondary ion mass spectrometry (SIMS) and Auger electron spectrometry (AES).

The most sensitive method to characterize the changes in the Schotty barrier for various experimental conditions seems to be the C-V measurement. Even when I-V and depth profiling measurements could not detect changes in the contact interface after high temperature annealing, C-V measurements can detect changes of the depletion layer depth and of the carrier concentration [13]. The stability of a silicide-GaAs gate depends mainly on the silicide film composition and on the processing techniques. For example, in W silicide, the optimum ratio of W to Si for a stable contact ranges from 5:2 to 3:1. When the film is too metal rich, the Si dopant in the GaAs interacts with W during high temperature annealing, and the doping concentration decreases. When the film is too Si rich, Si will diffuse into GaAs substrate, resulting in high doping near the interface [13]. From Table IV we also see that W-silicides prepared from composite targets are less stable than the films prepared from separate targets. The purity of targets and the cleanliness of the GaAs substrate surface are also critical in maintaining high temperature stability of Schottky contact [14]. These refractory metal silicides have been used successfully as a self-alinged gate in GaAs metal-semiconductor field effect transistor (MESFET) VLSI circuit [11, 12, 15].

VI. Epitaxial Silicides

Epitaxial silicides have many advantages (over polycrystalline silicides). Epitaxial silicides have better electrical, structural and morphological stability than polycystalline silicides. For instance, epitaxial Pd_2Si has a resistivity of 25 $\mu\Omega$ cm, while that of non-epitaxial films is between 30 to 35 $\mu\Omega$ cm [16]. Epitaxial PtSi and Pd_2Si are stable at temperatures at least 100°C higher than the non-epitaxial ones [17]. The interface between an epitaxial silicides and the Si substrate can be highly uniform, and with an amount of defects that can be controlled by the processing

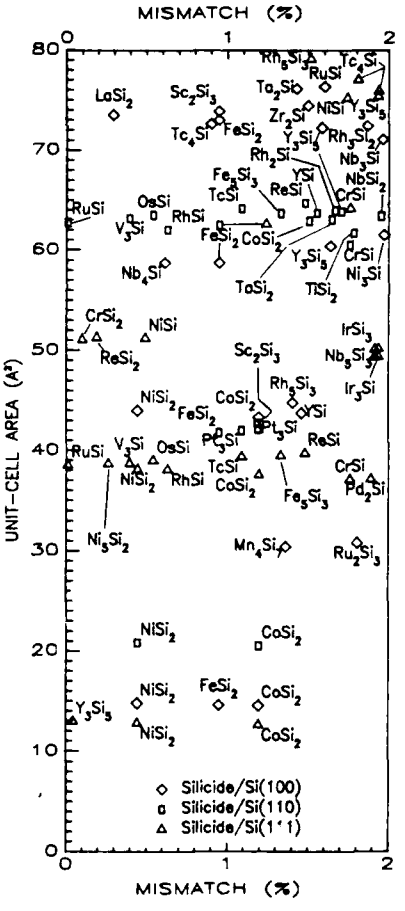


Fig. 3. Lattice matches of transition-metal silicides on Si (100), (110), (111). All the possible matches such that the mismatch is smaller than 2%, and the common unit-cell area is smaller than 80 A^2 are considered. These matches are characterized by mismatch and unit-cell area, and are shown here as diamonds for matches on Si (100), rectangles for matches on Si (110), and triangles for matches on Si (111). Only the better matches are labeled [18].

techniques. All these properties arise from the good crystalline perfection and the high degree of lattice matching with the Si substrate. By using such epitaxial films instead of polycrytalline silicides promise improved device performance and the reliability of Si integrated circuit. A main reason for studying epitaxial silicides is their potential for novel devices such as the metal base transistor (MBT) permeable base transistor (PBT), and infrared (IR) detectors as mentioned in Table I.

To master the growth of epitaxial films, many properties of the film and the substrate that influence the epitaxial growth have to be carefully investigated, among which material properties between the film and substrate are: lattice matching, thermal expansion coefficient, substrate orientation, interface chemistry, interfacial stress, diffusivity, and, thermodynamics. Successful epitaxial growth depends on a combination of these properties. A single criterion, such as a good lattice match, is neither sufficient nor necessary even though it is clearly useful as a guide. The lattice matches of silcides on Si is graphically summarized in Fig. 3 [18]. The processing parameters also influnece epitaxial growth. Such parameters include substrate cleaness, source purity, substrate heating, deposition rate, ambient, and annealing conditions. A partial summary of various growth conditions for silicide epitaxy on Si is given in Tables V to VIII. The literature on this subject is vast; the tables are meant to be representations, not exhaustive. The crystalline quality can be quantified by various techniques such as high resolution transmission electron micrscopy, double-crystal x-ray diffraction, or the channeling method of MeV He backscattering spectrometry. The tables list the channeling minimum yields (χmin). Cobalt and Ni disilicides have the lowest minimum yields among silicides which corresponds to the best epitaxial quality. For some non-single crystalline epitaxial silcides a typical grain size (G.S.) is cited.

Various growth methods are in use: For solid-phase epitaxy (SPE),the metal is deposited on Si at room temperature and annealed subsequently at elevated temperature to form the desired silicide phase. Reactive deposition epitaxy (RDE) is similar, except that the metal is deposited on heated Si substrate and further

Table VI. EPITAXIAL GROWTH OF REFRACTORY-METAL SILICIDES, PtSi AND Pd_2Si ON Si

Formation method	$CrSi_2$ on (111),(001) (011),Si	VSi_2 on (111) Si	$MoSi_2$ on (111),(001) (011) Si	$MoSi_2$,and Mo_3Si on (100) Si and Mo_3Si on (111)Si	$MoSi_2$ on (100) Si	WSi_2 on (111),(001) (011),Si	WSi_2 on (110)and (111) Si	$TiSi_2$ (111) Si	$MnSi_{1.7}$ on (111), (001) Si	$ZrSi_2$ on (111) and (001) Si
Metals e-gun deposited on a substrate held at 300-400°C. Two annealing steps of 300-500°C then 1000-1100°C promote epitaxial growth.	G.S.∿1-5μm ρ∿600μΩcm ϕ_B∿0.57eV [1]	G.S.∿1-4μm ρ∿50-55μΩcm ϕ_B∿0.54eV [1]	G.S.∿0.51 μm 100μΩcm ϕ_B∿0.55eV [1]			G.S.∿0.5 1 μm ρ∿70μΩcm ϕ_B∿0.65eV [1]		G.S.∿2-15 ρ∿10-25μΩ ϕ_B∿0.6eV [1]	G.S.∿ 0.5 μm ρ>1000 cm [1]	
VPE of a Si layer first, deposited from $SiCl_4$, then from $MoCl_5$ on Si substrate at 1000-1100°C.				thickness ∿20-30A G.S.∿0.1μm cylindrical growth [2]						
VPE of a Si layer first, deposited from $SiCl_4$,then Mo from MoO_2Cl_2 on a Si substrate at 1200°C.					thickness ∿20A cylindrical growth [2]					
VPE of a Si layer first, deposited from SiH_4+H_2, then W from WCl_6+H_2 on Si substrate at 1100-1200°C.							Film continuous whe. thickness >0.1μm [3]			
Metal e-gun deposited on a substrate held at 300-400°C.Two annealing steps of 300-500°C then 1000-1100°C. Ion mixing prior to annealing critical.										G.S.∿2-15μm ρ∿10-25 μΩcm ϕ_B∿0.06eV [1]

Table VI. EPITAXIAL GROWTH OF REFRACTORY-METAL SILICIDES, PtSi AND Pd_2Si ON Si (continued)

Formation method	$TaSi_2$ on (111)Si	$NbSi_2$ on (111)Si	$FeSi_2$ on (111)Si	Ru_2Si_3 on (1$\bar{1}$1)Si	RuSi on (111)Si	Ru_2Si on (1$\bar{1}$1)Si	Ir_2Si on (1$\bar{1}$1)Si	Rh_2Si_3 on (1$\bar{1}$1)Si	$ReSi_2$ on (111) and (001)Si	PtSi on (111)Si	Pd_2Si on (1$\bar{1}$1)Si
Metal e-gun deposited on heated substrate at 300-400°C, then furnace annealing.	$\rho\sim$35-45$\mu\Omega$cm $\phi_B\sim$0.59eV [1]	$\rho\sim$50$\mu\Omega$cm $\phi_B\sim$0.62eV [1]									
Metal e-gun deposited and ion mixed before annealing. Two annealing steps of 300-500°C then 1000-1100°C.			G.S.$\sim$7-10μm $\rho\sim$1000$\mu\Omega$cm [1]								
Metal deposited by electroless plating. Two annealing steps of 300°C then 900-1000°C.				G.S.$\sim$ 2-6μm [4]	G.S.$\sim$ 0.5-2μm [4]	G.S.$\sim$ 15μm [4]	G.S.$\sim$ 1μm [4]	G.S.$\sim$ 0.5μm [4]			
Metal deposited by e-gun and furnace annealing.									G.S.$\sim$ 0.5 - 0.7 μm [5]		
e-beam depositon of metal on Si and furnace annealing.										$X_{min}\sim$19-30% depending on temperature [6,7,8]	$X_{min}\sim$10% $\rho\sim$25$\mu\Omega$cm [6,7,8]
Room temperature 200 keV $5\times10^{16}/cm^2$ Ar irradiation through metal films on Si.											$X_{min}\sim$10% [10]

annealed afterwards to form a silicide layer. In molecular-beam epitaxy (MBE), metal and Si atoms are co-deposited on a heated Si substrate to form the silicide compound. In vapor-phase epitaxy (VPE), metal or Si atoms in compound vapor form are transported by a carrier gas to the Si substrate and thermally decomposed to form silicides. In the rapid thermal annealing (RTA) process, samples with a metal film deposited on Si are exposed to a halogen lamp or an electron beam for short periods of time (order of 10's of seconds) to form silicide compound. Electroless plating is another method by which to deposit a metal film before annealing. When metal ions in the 100 keV energy range are implanted into Si to a high dose, a subsequent thermal annealing cycle can lead to the formation of a buried silicide layer. Irradiation of a metal film deposited on a Si substrate by energetic ions (ion mixing) can also induce an epitaxial growth.

Certain silicides seem to form better by some techniques than by others. Some rare earth silicides form a smoother films with less pitted surfaces by e-beam heating than by solid-phase epitaxy [19]. The main methods used for the synthesis of epitaxial silicides have been MBE, SPE, and RDE techniques. By some, MBE is the favored technique for high quality epitaxial films with low pinhole density and defects [20] on the rationale that for MBE the total change in temperature during processing is least. However, there are instances [21 and 22], where a high quality heterostructure of silicide/Si has been successfully grown by SPE method. Especially, there is a method by which epitaxial silicides can be grown in a superior quality, which is known as the "template technique" [23]. In this technique, a thin, uniform, single crystal seed layer of silicide is grown by depositing the metal at moderate temperature in ultra high vacuum. On top of this seed layer, a thick layer of silicide is further grown by metal deposition at elevated temperatures. By combining other deposition techniques in various ways, a large number of possibilities exist, not all of which have yet been published.

The cleaning procedures applied to the Si substrate prior to the actual epitaxial growth is very critical step. Widely used and successful

is the so-called "Shiraki" procedure [24] which consists of chemical oxidation-etching cycles that leave a very thin protective SiO_x film. This film is evaporated at about 900°C just before the growth begins. Another popular technique of surface cleaning is to ion sputter the Si surface with low energy Si or Ar ions and then anneal it [25]. The goal of these techniques is to remove a surface layer of Si that may contain impurities, thus exposing an anatomically clean surface for the epitaxial growth.

$CoSi_2$ and $NiSi_2$ epitaxial films have two possible orientations (A or B types) when grown on Si(111). In type A, the Si atoms in the silicide film extend the stacking order of the Si {111} substrate without fault because their unit cell is that of CaF_2, not diamond; in type B, the Si stacking order is mirror-imaged at the interface, which amounts to a rotation of the Si lattice by 180° about the surface normal. What type of silicide grows can be controlled by template method. The rate of metal deposition on Si to form the template also has influence in determing the orientation type of epitaxial silicide [26]. The Schottky barrier height on n-Si has been reported to vary with the type (A or B) silicides, [27]; another reference [28] attributes this effect to defects and impurities (see Table VIIa). Among other numerous current investigations, the reduction of pinhole densities [29] and the understanding of the kinetics of epitaxial silicide formation [30] are hotly pursued.

Table VIII lists silicides that have been formed by high dose implantations of metal ions into Si. Upon post annealing, layers of silicides can be formed below the Si surface. $CoSi_2$ layers of excellent epitaxial quality can be obtained [31]. Buried epitaxial silicide layers could bring heterostructure devices (such as the metal base transistor (MBT), or the

Table VII.

A) EPITAXIAL FORMATION OF $NiSi_2$ ON {111} Si

Formation methods	Properties	References
Template growth method; deposit 1-6x10^{15} Ni/cm^2 on Si substrate at room temperature in UHV condition and in-situ anneal it to 450-550°C. Deposit further Ni at 500-755°C to form final $NiSi_2$ layer.	Single crystal B-type Xmin 3%. Deposition rate of initial Ni also influences the type of $NiSi_2$ (A or B) $\phi_B \sim 0.76$-0.79eV	1,2,3,4,5
Template growth method: deposit 12-17x10^{15} Ni/cm^2 on Si substrate at room temperature in UHV and anneal it to 450-550°C. Deposit further Ni at 500-775°C to form final $NiSi_2$ layer	Single crystal A-type. Xmin$\sim$3% ϕ_B 0.62-0.65eV	1,2,3,4,5
Template growth method to grow either single cyrstal A or B type $NiSi_2$	A or B type with significant amount of defects and impurities in the interface $-\phi_B \sim 0.66$eV. A or B type with near perfect interface $\phi_B \sim 0.78$eV	6
Template growth method with annealing temperature of 250°C to grow NiSi on Si	Sample with significant amount of defects and impurities in the interface $\phi_B \sim 0.66$eV. Sample with near perfect interface $\phi_B \sim 0.78$eV.	6
SPE with annealing temperature of 400°C	B-type $NiSi_2$	7
SPE with annealing temperature of 700-800°C	Xmin 4% $\rho \sim 25\ \mu\Omega$cm	8
MBE growth at substrate temperature of 400°C	Perfect B-type $NiSi_2$	7
Ni deposited on Si and annealed by Xe arc lamp RTA in N_2 ambient at temperatures 700-800°C	Xmin$\sim$4.6%	9
B) EPITAXIAL FORMATION OF $CoSi_2$ ON {111} Si		
SPE with annealing temperature of 650-750°C	Xmin$\sim$ 3.6%	10
SPE with annealing temperature of <100°C	Co_2Si and epitaxial CoSi are formed	11

Table VII.

B) EPITAXIAL FORMATION OF $CoSi_2$ ON {111} Si (continued)

Formation methods	Properties	References
SPE with annealing temperature of 570°C	$CoSi_2$ with 80% B-type	11
SPE with annealing temperature of 900-1000°C	$\chi_{min} \sim 3\%$ 15 cm	8
SPE with annealing temperature of 1050°C	$CoSi_2$ + Si or CoSi formed $\chi_{min} \sim 5\%$	8
Template growth method: deposit 220A of Co at room temperature in UHV and anneal it at 975°C. Deposit further Co at 975°C to from final $CoSi_2$ layer	B-type, $\chi_{min} \sim 4\%$. Growth is planar. Flat interfaces and uniform thickness	1
MBE growth at substrate temperature of 600-650°C	B-type, $\chi_{min} \sim 2\%$. Free of grain boundaries	12
MBE growth at substrate temperature of 580°C	$\chi_{min} \sim 3.6\%$	10
RDE - Co was deposited on Si at 580°C	$\chi_{min} \sim 3.6\%$	10
RDE - Co deposited on Si at 250°C. Samples further annealed at 800-1000°C	film is 92% B-type and 8% A type	13
RDE - Co deposited on a Si at 100°C. Samples are further annealed at 550 and 625°C	$\rho \sim 15 \mu\Omega$cm (200K) for 80Å film $\rho \sim 117$ cm (200K) for 12A film 12Å is B-type only. Films of above 42A consist of both A and B-types	14
Co deposited on Si and annealed by Xe arc lamp RTA in N_2 ambient at 750°C	$\chi_{min} \sim 5.7\%$ $\rho \sim 18 \mu\Omega$cm	9
Porous Si<111> substrate formed firstly. Secondly, surface a buffer layer of epitaxial Si of 150A was grown on the porous surface. Lastly, epitaxial $CoSi_2$ grown by MBE method on Si at 500°C.	$\chi_{min} \sim 12\%$ for 240Å $CoSi_2$. As the film grows above 500A the cyrstallinity improves	16
C) EPITAXIAL FORMATION OF $NiSi_2$ ON {100} Si		
Use template growth method; deposit 10-30A Ni at room temperature and in-situ anneal the sample at 450-550°C. Deposit more Ni at 650°C to form final $NiSi_2$ layer	$\chi_{min} \sim 4.5\%$. Same orientation as the Si substrate	15
D) EPITAXIAL FORMATION OF $CoSi_2$ ON {100} Si		
SPE with Annealing temperature of 900°C	$\chi_{min} > 45\%$	8
RDE - Co was deposited on a substrate at 230°C. The samples are annealed at 900°C afterwards	G. S. ~ 0.5-1.0 μm	5

*χ_{min} - Channeling yield minimum
ϕ_B - Schottky barrier height
G.S. - Grain size
SPE - Solid phase epitaxy
MBE - Molecular-beam epitaxy
RDE - Reactive deposition epitaxy

Table VIII. SILICIDE FORMATION BY HIGH DOSE IMPLANTATIONS OF METAL IN Si

Implanted* Metal, Si orientation, Silicide	Energy and Dose	Implantation Temperature	Post annealing temperature	Quality	Ref.
Co in 100 Si, 1100A of $CoSi_2$ was formed with 600A Si overlayer	200 keV $2\text{-}3\times10^{17}/cm^2$	350°C	600-1000°C and 1000°C	$\chi_{min}\sim10\%$ T 1.55 k RR 15.5 $\rho\sim12.4\mu\Omega cm$ R.T	1
Co into {111} Si, 1100A of $CoSi_2$ formed with 600A Si overlayer	same as above	350-450°C	same as above	χ_{min} 6.4% Tc 1.49 k RR 7.8	1
Co, {100} Si, $CoSi_2$	350 keV $1\text{-}6\times10^{17}/cm^2$	450°C	1000°C	χ_{min}-15% residual strain: $\epsilon\perp$ =-1% $\epsilon\parallel$ =0.6%	2
Cr, {100} Si, $CrSi_2$	100keV, $5\times10^{17}/cm^2$	R.T	550°C		3
Fe, {100}Si, $FeSi_2$	100 keV $6\times10^{17}/cm^2$	R.T.	400°C		3
Ni {100}Si, $NiSi_2$	100 keV $6\times10^{17}/cm^2$	R.T.	400°C		3
Ni, {100}Si, NiSi	150 keV $6\times10^{17}/cm^2$	350°C			3
Co, {100}Si, CoSi	100 keV $1.6\times10^{18}/cm^2$	R.T.			3
Ti, {111} Si, 600A $TiSi_2$ formed with 900A Si overlayer	120 or 170keV	T>450°C	600°C	residual damage from implantation removed after 100°C annealing	
V,Cr,Nb,Ti and Mn, {100}Si, VSi_2, $CrSi_2$, $NbSi_2$, Ti_5Si_3, and Mn_5Si_3	150 keV $1\text{-}5\times10^{17}/cm^2$	350°C			5
Fe,Co,Ni into {100}Si, FeSi, CoSi, $NiSi_2$		150 keV 350°C			5

permeable base transistor (PBT) close to reality. Research in this area is being pushed aggressively.

The MBT and PBT have high current drive capability and are of vertical devices in structure, as is the conventional Si bipolar junction transistor (BJT). The MBT or PBT equipped with a metallic epitaxial silicide promise advantages over the BJT in terms of speed. Both structures have been succesfully fabricated [20, 32, 33]. Less than 1% of current transmission was observed in MBT. The reason is believed to be inelastic electron scattering and the quantum mechanical reflection at the base-collector interface [34]. In concept, the PBT is a vertical FET. Channels of Si pass throughout the silicide layer and carry current that is controlled by the field established with the metallic grid acting as the gate [32]. By varying the Si channel width the current transmission was obtained [32]. The high frequency capabilities of the PBT has been tested.

Epitaxial silicides on Si also have applications as IR detectors whose operation is based on the Schotty-barrier nature of a silicide/silicon contact and the photoemission of carriers from the silicide to Si [35, 36]. The barrier height determines the cut-off wavelength. There is also a proposal to use semiconductor silicides with direct band gaps for opto-electronic devices to detect and generate light [37]. $FeSi_2$ seems to have a direct band-gap of 0.87eV; $CrSi_2$, $MnSi_{1.7}$, and

$IrSi_{1.7}$ are semiconducting silicides too. With silicide detectors, direct band gap semiconducting silicides, and the signal-processing circuitry of a Si chip, the major elements for a high-speed integrated opto-electronic chip fully based on Si exist in principle. Much research presently goes into the marrying of GaAs opto-devices with a Si based chip. Combining GaAs and Si technologies is proving difficult. By comparison, the development of silicide based opto-electronic devices on Si has the advantage of cenceptual simplicity that is now being recognized.

VI. Acknowledgments

The authors would like to acknowledge Rose Pieters-Emerick for preparation of the manuscript. This paper was written under partial support by National Science Foundation the Semiconductor Research Corporation under contract #87-SJ-100. A grant by Intel Corporation is also gratefully acknowledged.

Text References

1. Nicolet, M-A., and S.S. Lau, in VLSI ELectronics Microstructure Science, Vol. 6, N.G. Einspruch and G.B. Larrabee, eds. Academic Press, New York, 329 (1983).
2. Murarka, S.P., Silicides for VLSI Applications, Academic Press New York, (1983).
3. Chow, T.P., and A. J. Steckl, in VLSI Electronics: Microstrucuture Science, Vol. 9., N.G. Einspruch, ed. Academic Press, New York (1981).
4. d'Heurle, F.M., and P. Gas, J. Mat. Res. 1, 205 (1986).
5. Bucher, E., S. Schulz, M.D. Lux-Steiner, P. Munz, U.Gubler and F. Grueter, Appl. Phys. A40, 71 (1986).
6. Haken, R.J., J. Vac. Sci. Technol. B3, 1657 (1985).
7. Koh, Y., F. Chien and M. Vora, J. Vac. Sci. Technol. B3, 1715 (1985).
8. Ho, V.W., and D. Poulin, J. Vac. Sci. Technol. A5, 1396 (1987).
9. Murarka, S.P., J. Vac. Sci. Technol. B4, 1325 (1986).
10. Yokoyama, N., T. Ohnishi, K. Odani, H. Onodera, and M. Abe, IEDM Tech. Dig., 80 (1981).
11. Tseng, W.F., B. Zhang, D. Scott, S.S. Lau, A.Christou and B.R. Wilkins IEEE Electr. Dev. Lett. EDL-4, 207 (1983).
12. Cirillo, Jr., N.C., H.K. Chung, P.J. Vold, M.K. Hibbs-Brenner and A.M. Fraasch, J. Vac. Sci. Technol. B3, 1680 (1985).
13. Jackson, T.N., and J.F. DeGelormo, J. Vac. Sci. Technol. B3, 1676 (1985).
14. Zhongde, Z., N.W. Cheung, Z.J. Lemnios, M.D. Strathman and J.B. Stimmell. J. Vac. Sci. Technol. B4, 1398 (1986).
15. Tseng, W.F., and A. Christou, Electr. Lett 19, 330 (1983).
16. Saitoh, S., H. Ishiwara, T.Asano and S. Furukawa, Jpn. J. Appl. Phys. 20, 1649 (1981).
17. Ishiwara, H., Proc. of Symp. on Thin Films Interfaces and Interactions, Eds. J.E.E. Baglin and J.M. Poate, Electrochem. Soc, New York, 159 (1980).
18. Zur, A., T.C. McGill and M-A. Nicolet, J. Appl. Phys. 57, 600 (1985).
19. Knapp, J.A., and S.T. Picraux, Appl. Phys. Lett. 48, 466 (1986).
20. Wang, K.L., and Y.C. Kao, Mat. Res. Soc. Symp. Proc. 67, 235 (1986).
21. D'Avitaya, F.A., J.A. Chroboczek, C. D'Anterroches, G. Glastre, Y. Campidelli and E. Rosencher, J. Crystal Growth 81, 463 (1987).
22. Tung, R.T., and J.M. Gibson, Mat. Res. Soc. Symp. Proc. 14, 395 (1983).
23. Tung, R.T., J.M. Gibson and J.M. Poate, Appl. Phys. Lett. 888 (1983).
24. Ishizaka, A., K. Nakagawa, Y. Shiraki: Proc. of the Symp. on MBE and Clean Surface Technologies, (The Japan Society of Appl. Phys., Tokyo, 183 (1982).
25. Wang, K.L., Solid State Tech. 28(10), 137 (1985).
26. Tung, R.T., J. Vac. Sci. Technol. A5, 1840 (1987).
27. Tung, R.T., Phys. Rev. Lett. 52, 461 (1984).
28. Lieh, M., P.E. Schmid, F.K. LeGoues, and P.S. Ho, Phys. Rev. Lett. 54, 2139 (1985).
29. e.g. Tung, R.T., and J.L. Batstone, Appl. Phys. Lett. 52, 648 (1988).
30. e.g. Veuillen, J.Y., J. Derrien, P.A. Badoz, E. Rosencher and C. d'Anterroches, Appl. Phys. Lett 51, 1448 (1987).
31. White, A.E., K.T. Short, R.C. Dyres, J.P. Garno and J.M. Gibson, Appl. Phys. Lett. 50, 95 (1987).
32. Hensel, J.C., A.F.J. Levi, R.T. Tung and J.M. Gibson, Appl. Phys. Lett. 47, 151 (1985); R.T. Tung, A.F.J. Levi and J.M. Gibson, Appl. Phys. Lett 48, 635 (1986).
33. Rosencher, E., S. Delage, Y. Campidelli and F.A. D'Avitaya, Electron. Lett. 20, 762 (1984); E. Rosencher, G. Glastre, G. Vincent, A. Vareille and F.A. D'Avitaya, Electon. Lett. 22, 699 (1986).
34. Levi, A.F.J., R.T. Tung, J.L. Batstone and M. Amzlowar, Mat. Res. Soc. Symp. Proc. 102, (1988).
35. Kosonocky, W.F., H.G. Erhardt, G. Meray, F.V. Shallcross, H. Elabd, M.J. Cantella, J. Klein, L.H. Skolnik, B.R. Capone, R.W. Taylor, W. Ewing, F.D. Shepherd and S.A. Rooslid, SPIE Proc. 22, 69 (1980).
36. Harrison, T.R., A.M. Johnson, P.K. Tien and A.H. Dayem, Appl. Phys. Lett. 41, 734 (1982).
37. Bost, M.C., and J.E. Mahan, J. Vac. Sci. Technol. B4, 1336 (1986).

Table I References

1. Murarka, S.P., J. Vac. Sci. Technol. B4, 1325 (1986).
2. Rosencher, E., S. Delage, Y. Campidelli, F. A. D'Avitaya, EL 20, 762 (1984).
3. Ishizaka, A., and Y. Shiraki, Jpn. J. Appl. Phys. 23, (1984) L499.

4. Tung, R.T., A.F.J. Levi and J.M. Gibson, Appl. Phys. Lett. 48, 635 (1986).
5. White, A.E., K.T. Short, R.C. Dynes, J.P. Garno, and J.M. Gibson, Appl. Phys. Lett. 50, 95 (1987).
6. Wang, K.L., G.P. Li, IEEE EDL-4, 444 (1983).
7. Wang, K.L., Y.C. Cao, Mat. Res. Soc. Symp. Proc. Vol. 67, 235 (1986).
8. Harrison, T.R., A.M. Johnson, P.K. Tien, and A.H. Dayem, Appl. Phys. Lett. 41, 734 (1982).
9. Kosonocky, W.F., H.G. Erhardt, G. Meray, F.V. Shallcross, H. Elabd, M.J. Cantella, J. Klein, L.H. Skolnik, B.R. Capone, R.W. Taylor, W. Ewing, F.D. Shepherd and S.A. Rooslid, SPIE Proc. Vol. 22, 69 (1980).
10. Bost, M.C., and J.E. Mahan, J. Vac. Sci. Technol. B6, 1336 (1986).
11. Fauthauer, R.W., P.J. Grunthaner, T.L. Lin, K.T. Chang, J.H. Mazur and D.N. Jamieson, to be published in J. Vac. Sci. Technol. B (1988).
12. Tokushima, T., I. Nishida, K. Sakata and T. Sakata, J. Mat. Sci. 4, 987 (1969).
13. Ware, R.M., D.J. McNeill, Proc. IEEE 111, 178 (1964).
14. Kwok, S.P., J. Vac. Sci. Technol. B4, 1383 (1986).
15. Cirillo, N.C., Jr., H. K, Chung, P.J. Vold, M.K. Hibbs-Brenner and A.M. Fraasch, J. Vac. Sci Technol. B3, 1680 (1985).
16. Tseng, W.F., B. Zhang, D. Scott, S.S. Lau, A. Christou, B.R. Wilkins, IEEE, Electr. Dev. Lett. EDL-4, 207 (1983).
17. Tu, K.N., K.Y. Ahan and S.R. Herd, Appl. Phys. Lett, 39, 927 (1981).
18. Bartlett, R.W., P.R. Gage and P.A. Larssen, TRANS. AIME 230, 1528 (1964).
19. Bartlett, R.W., TRANS. AIME 236, 1230 (1966).

Table II References

1. Chow, T.P., and A. J. Steckl, IEEE, Trans Electr. Dev. Vol. Ed-30 1480 (1983).
2. Chow, T.P., and A. J. Steckl, in VLSI Electronics: Microstructure Science, Edited by N.G. Einspruch, Vol. 9, Academic Press, New York, (1981).
3. Sakiyama, K., Y. Yamauchi and K. Matsuda, J. Vac. Sci. Technol. B3, 1685 (1985).

Table V References

1. Knapp, J.A., and S.T. Picraux, Appl. Phys. Lett. 48, 466 (1986).
2. Knapp, J.A., and S.T. Picraux, Mat. Res. Soc. Symp. Proc. 54, 261 (1986).
3. Furvitch, M., A.F.J. Levi, R.T. Tung and S. Nakahara, Appl. Phys. Lett. 51, 311(1987).

Table VI References

1. Chen, L.J., H.C. Cheng and W.T. Lim, Mat. Res. Soc. Symp. Proc. 54, 245 (1986).
2. Casey, J.J., R.R. Verderber and R.R. Garnache, J. Electrochem. Soc. 114, 201 (1967).
3. Hashimoto, N., and Y. Koga, J. Electrochem.
4. Chang, Y.S., J.J. Chu, and L.J. Chen., Mat. Res. Soc. Symp. Pro . 54, 57 (1986).
5. Chen, J.J., L.J. Chen and K.N. Tu, J. Appl. Phys. 62, 461 (1987).
6. Ishiwara, H., Proc. of Symp. on Thin Film Interfaces and Intereactions, Eds., J.E.E. Baglin and J.M. Poate Electrochem. Soc., New York 159 (1980).
7. Ishiwara, H., K. Hikosaka and S. Furukawa, J. Appl. Phys. 50, 5302 (1979).
8. Tung, R.T., J.M. Poate, J.C. Bean, J.M. Gibson and D.C. Jacobson, Thin Solid Films, 93 (1982) 77.
9. Saitoh, S., H. Ishiwara, T. Asano and S. Furukawa, Jpn. J. Appl. Phys., 22 1118 (1981).
10. Ishiwara, K., and N. Kuzuta, Appl. Phys. Lett. 37, 641 (1980).

Table VII References

1. Tung, R.T., J.M. Gibson and J.M. Poate, Appl. Phys. Lett. 42, 888 (1983).
2. Tung, R.T.,J. Vac. Sci. Technol. A5, 1840 (1987).
3. Tung, R.T., Phys. Rev. Lett., 52, 461 (1984).
4. Hauenstein, R.J., T.E. Schlesinger and T.C. McGill, Appl. Phys. Lett. 47, 853 (1985).
5. Ospelt, M., J. Henz, L. Flepp and H. von Kanel, Appl. Phys. Lett 52, 227 (1988).
6. Liehr, M., P.E. Schmid, F.K. LeGoues and P.S. Ho, Phys. Rev. Lett. 54, 2139 (1985).
7. von Kanel, H., T. Graf, J. Henz, M. Ospelt and P. Wachten, J. Crystal Growth 81, 470 (1987).
8. Saitoh, S., H. Ishiwara, T. Asano and S. Furukawa, Jpn. J. Appl. Phys. 20, 1649 (1981).
9. Chevallier, J., and A.N. Larsen, Appl. Phys. A 39, 141 (1986).
10. Hamdi, A.H., M-A. Nicolet, Y.C. Kao, M. Tejunani and K.L Wang, Mat. Res. Soc. Symp. Proc. 41, 355 (1985).
11. D'Anterroches, C., Surface Science 168, 751 (1986).
12. Tung, R.T., J.C. Bean, J.M. Gibson, J.M. Poate and D.C. Jacobson, Appl. Phys. Lett. 40, 684 (1982).
13. Bulle-Lieuwama, C.W.T., A.H. Von Omnen and J. Hornstra, Mat. Res. Soc. Symp. Proc. 101, (1988).
14. Phillips, J.M., J.L. Botstone, J.C. Hensel and M. Cerullo, Appl. Phys. Lett. 51, 1895 (1987).
15. Tung, R.T., J.M. Gibson and J.M. Poate, Phys. Rev. Lett. 50, 429 (1983).
16. Kao, Y.-C., Ph.D. Thesis, Electrical Engineering, UCLA (1987).

Table VIII References

1. White, A.E., K.T. Short, R.C. Dynes, J.P. Garno and J.M. Gibson APL 50, 95 (1987); Mat. Res. Soc. Symp. Proc. 74, 481 (1987).
2. Barbour, J.C., S.T. Picraux and B.L. Boyle, Mat.Res. Soc. Proc. 101, (1988).
3. Sanchez, F.H., F. Namavar, J.I. Budnick, A.

Fasihuddin and H.C. Hayden, Mat. Res. Soc. Proc. 51, 439 (1986).
4. Madakson, D., G.J. Clark, F. K. Legoues, F.M. d'Heurle and J.E.E. Baglin, Mat. Res. Soc. Symp. Proc. 101, (1988).
5. Namavar, F., F.H. Sanchez, J.I. Budnick, A.H. FasiHuddin and H.C. Hayden, Mat. Res. Soc. Symp. Proc. 74, 487 (1987).

ELECTROMIGRATION

Ilan A. Blech
Zoran Corporation
3450 Central Expressway
Santa Clara, California 95051, USA

ABSTRACT

The phenomena of electromigration, material movement resulting from electrical currents, has been recognized in the past 23 years as one of the failure causes in microelectronic devices(1). This paper is a short review of the various failure modes due to electromigration. These include open electrical paths in conductor films, opens or high resistance in metal/semiconductor contacts, opens in intermetal vias and metal-metal short. The basic equations governing the electromigration mass flow in bulk conductors and its extension to thin films are shown, important physical attributes contributing to this flow are highlighted. Other important driving forces for mass flow in thin films are shown to be stresses and concentration gradients. The electromigration failure rate measurement techniques are also shown. Better electromigration simulations and failure predictions are emerging recently, these will be discussed. Finally, methods of electromigration reduction are presented, these methods include overcoats, alloying, refractories and laminates.

IN THE MIDDLE OF the 1960s there was a strong activity at Fairchild(1) and IBM(2) to understand the reasons for failures observed in the thin aluminum conductor films on transistors and integrated circuits. The failures manifested themselves as a series of voids in the conductors causing open circuits. It was found that the voids developed in the conductors as a result of passage of a direct current in the film. At that time there was no limit set for the allowed current density and typical current densities reached several times $10^5 A/cm^2$. At those current densities there is an appreciable atom flow due to the electrical currents. This phenomenon, known as electromigration, was identified as responsible for void formation and eventual failure. This paper presents a short overview of this phenomenon and the methods used to reduce the failure rate.

DEFINITION

By electromigration we mean the movement of material in a solid conductor due to the flow of electrical currents. The material flow can be in the direction of the electron flow or opposite to it. Most materials that are used as conductors in microelectronics migrate in the direction of the electron flow.

MASS FLOW EQUATION

In the early 1960's Huntington and coworkers(3) were engaged in measurement of the bulk electromigration mass flow. These measurements were done by observing the migration of fine scratches on the surface of wires or hollow cylinders, as a result of high electrical current densities at elevated temperatures. The equation used for determining the mass flow was,

$$\text{Atomic flux} = \text{Force} \times \text{Mobility} \times \text{Ion Density} \quad (1)$$

where,

$$\text{Force} = eZ^*\rho j \quad (2)$$

$$\text{Mobility} = D/kT \quad (3)$$

$$\text{Ion Density} = N \quad (4)$$

where,

J - Atom flux.
N - Ion density.
D - Diffusion coefficient.
$D=D_0\exp(-Q/kT)$ D_0 - pre-exponential factor,
Q - Activation energy.
k - Boltzmann's constant.
T - Absolute temperature.
e - Electronic charge.
Z^*- Effective charge.
ρ - Electrical resistivity.
j - Electrical current density.

Substituting Eqs (2), (3) and (4) into Eq. (1) leads to:

$$J=(ND/kT)eZ^*\rho j \quad (5)$$

The average drift velocity, V, is,

$$V = J/N \quad (6)$$

Grone(4) measured the bulk aluminum electromigration rate at the temperature range of 460 to 640°C at current densities of 10^4 A/cm^2. His results extrapolated to 225°C at 10^6 A/cm^2 indicate a very low drift velocity of 0.01A/S.

Thin film electromigration rates are thought to obey Eq. (5) with the exception that the mobility is the grain boundary mobility rather than the bulk(5). Since the grain boundary activation energy, is lower than that of the bulk, it follows that the bulk electromigration rate declines faster with decreasing temperatures than that of the thin film. As an example, the extrapolated electromigration drift velocities for both aluminum bulk and thin films are given in Table 1. The thin film drift velocity is dependant on its grain size, so that the results in Table 1 are not universal but only indicative of expected thin film drift velocity values.

Table 1. Electromigration drift velocities for aluminum(4,5).

Temperature	Bulk	Thin Film
630°C	1500A/S	33 A/S
225°C	0.01A/S	0.6A/S

It can be seen that while the bulk electromigration drift velocity is 50 times faster than that of thin films at 630°C, it is 60 slower at 225°C. Electromigration drift velocities(6) and activation energy measurements(7) were summarized in the literature.

ELECTROMIGRATION FAILURE MODES

The fact that material moves under the influence of electrical currents is not by itself a cause for failure of the conductors. The *divergencies* in such flows are the prime cause for failure. If a certain area of the conductor receives more material than is depleted from it, than a net accumulation of material occurs with resulting formation of hillocks or whiskers which eventually can cause a short circuit. If, on the other hand, material is depleted faster that replenished, a net material loss occurs which results in void formation and eventual open conductor line.

A number of failure modes are schematically shown in Figure 1. This figure, which shows a cross-section of a two level metal integrated circuit, depicts 6 different failure modes due to electromigration. Open lines and whiskers have already been discussed here, other failures are metal1/metal2 shorts, open vias and contact problems. A metal1/metal2 short can occur when material is accumulating on metal1 causing the inter-metal dielectric to fracture or lift. The metal can then extrude through the crack to form a short(8). Open vias can occur if a refractory (low electromigration rate) barrier is used between metal1 and metal2. The conductor metal can drift away from the via which connects metal1 and metal2 leaving a void behind. Contact problems related to electromigration have also been reported (9-11). In essence, the problems are related to silicon electromigrating through aluminum metallization at a higher rate than the aluminum itself. The silicon is depleted on one side (where electrons flow out of the contacts) causing deep holes (spikes) in the silicon, and on the other side (where the electrons flow into contacts) the silicon forms a high resistance deposit. Both occurrences are detrimental, in the first case the spikes can traverse pn junctions and cause high leakage currents while in the second case the deposit causes high electrical resistance contacts.

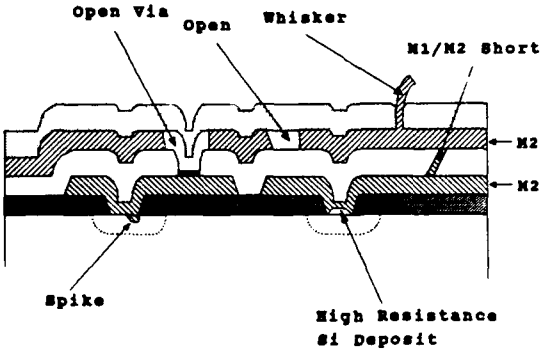


Figure 1. Schematic cross section of a circuit showing six various failure modes due to electromigration.

There are several causes for a divergence in mass flow all related to changes in the mobility. There can be changes in *temperature*, in *composition* and *structure* along the conductors. Figure 2 shows schematically the effects of these parameters. The temperature change is the easiest to comprehend - if the temperature rises along the direction of the atom flow in a conductor, it is clear that material is being moved out of the conductor faster (higher temperature) as being supplied and such a conductor will therefore develop voids in it. The opposite will occur when the temperature decreases in the direction of atom flow and hillocks or whiskers will form there. Temperature gradients can form in a variety of places in microelectronic devices, at contacts, steps or width changes.

By composition changes we mean a change in the conductor material such as a conductor which changes locally from tungsten to aluminum. Such a change in material will cause a dramatic change in electromigration flow rates, in this example the tungsten migrates very slowly while the aluminum migrates fast. Voids will occur in the aluminum where the electrons (and the atoms) flow away from the tungsten into the aluminum(12).

Finally, the microstructure of the conductor film affects the void and hillock formation. Figure 2 shows this schematically, where a void is seen to develop in grain boundary triple points. A void forms where the outgoing flux is higher than the incoming and a hillock

forms where the opposite occurs. Another structure change was reported recently as being responsible for aluminum voiding or accumulation; the finer grain structure which develops in aluminum over contact areas creates a flux gradient which leads to an increase failure rate at the contacts(13).

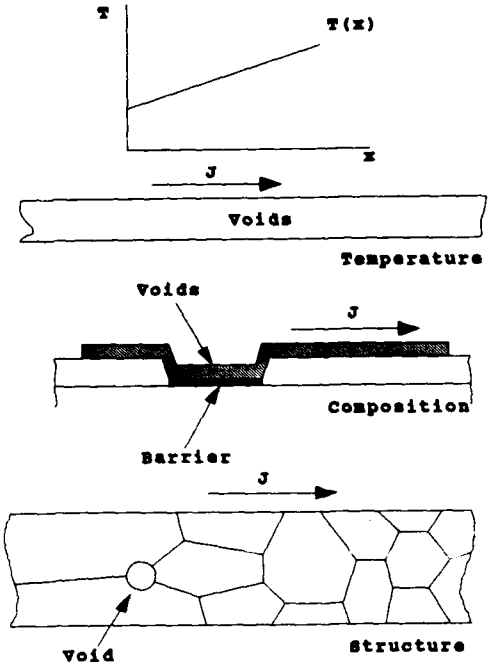


Figure 2. Schematic diagram showing the deleterious effects of temperature, composition and structural gradients.

FLUX BACKFLOWS

The concept of backflows during electromigration was suggested a number of years ago(5). It simply suggests that when the electromigration forces an atom flow, the atoms can congest at certain areas causing a local increase of compressive stress leading to a backflow of atoms. Similarly, when an alloy conductor is under electrical current, the faster moving constituent will accumulate and create a concentration driven backflow(14). It is these backflow that will alter the net electromigration flow. Figure 3 shows schematically a short and long conductor carrying the same current density, in both cases the electrons flow from left to right and a compressive stress is created on the right side of the conductors. The maximum compressive stress, S, is that stress needed to form a hillock or perhaps, in the case of an encased conductor, the stress necessary to fracture the encasement. The backflow will be proportional to the stress gradient (S/L). The forward flow due to electromigration is proportional to the current density, j, according to Eq. 5 and can be written as Cj, where C is a proportionality constant. The net flow will therefore be proportional to Cj-S/L. The stress gradient will be larger in the shorter conductor (L being smaller) resulting in a smaller net atom flow in the shorter conductor. In fact, if the conductor will be short enough, there will be no flow at all (when the stress backflow equals the electromigration flow). The length at which the flows equal was termed the critical length(5), L_c, and from Figure 3 it can be seen that zero flow occurs when jL_c=constant.

We expect from the foregoing that the mass flow will not be proportional to the current density but will decrease dramatically when the conductor length approaches a critical length. The critical length was actually measured and found to be a function of temperature and grain size(5).

The concept of stress backflow has been confirmed by actual stress measurements along current carrying aluminum conductors(15,16). A stress gradient proportional to the conductor length was seen to be created by the electrical currents.

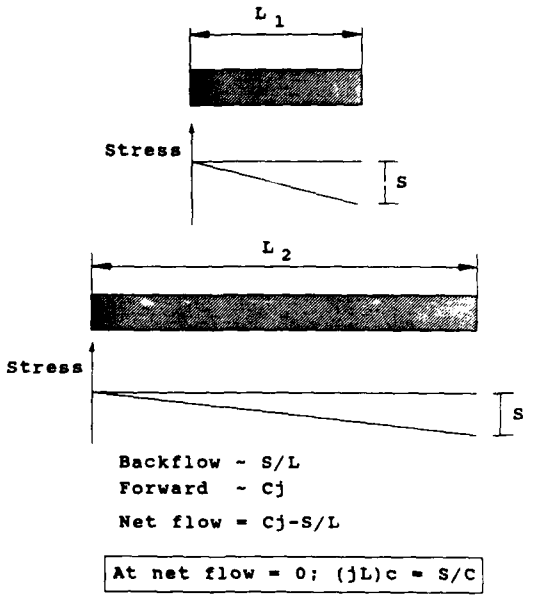


Figure 3. Schematic diagram showing that short conductors have a larger stress gradient and therefore a smaller net flow of atoms during electromigration.

The idea of stress backflow was extended from confined conductors to infinitely long conductors(17-19). The critical length is now replaced from the actual conductor length to a void-hillock length. Figure 4 shows, schematically, a portion of a long conductor with several voids and hillocks forming on it. The backflows occur between the hillocks and holes, reducing the total electromigration damage, and again causing the lifetime to increase nonlinearly at low current densities.

The backflows have the tendency of reducing the deleterious effects of some of the "dangerous geometries" in integrated circuits. Figure 5 shows these effects, as an example, the thinning of metal at steps creates a higher electromigration flux which in turn will create a large backflow since the local stress gradient is high, reducing the possibility of failure at the step. The stress gradient is not the only factor governing the lifetime of the thinned portion over steps, the temperature gradients and the structural gradients are also very important. It

should be emphasized here that the backflow effects become important at the low current densities actually used in the circuits. At high current densities, normally used in accelerated lifetests, the backflow effects are usually not significant. The lifetimes extrapolated from the high current density accelerated measurements are actually pessimistic since they do not take into account the beneficial effects of the backflows. The actual lifetimes are expected to be significantly higher when the current density is close to the current at which the forward electromigration flow is balanced by the backflow.

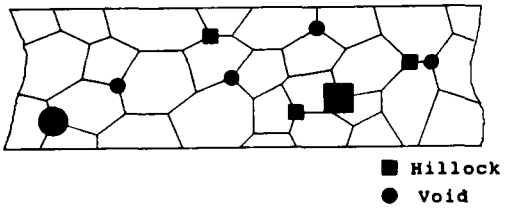


Figure 4. Schematic diagram showing hillocks and void formation on a polycrystalline conductor. An atom backflow will occur from the hillocks (excess atoms, compression) to the voids reducing the electromigration flow and damage.

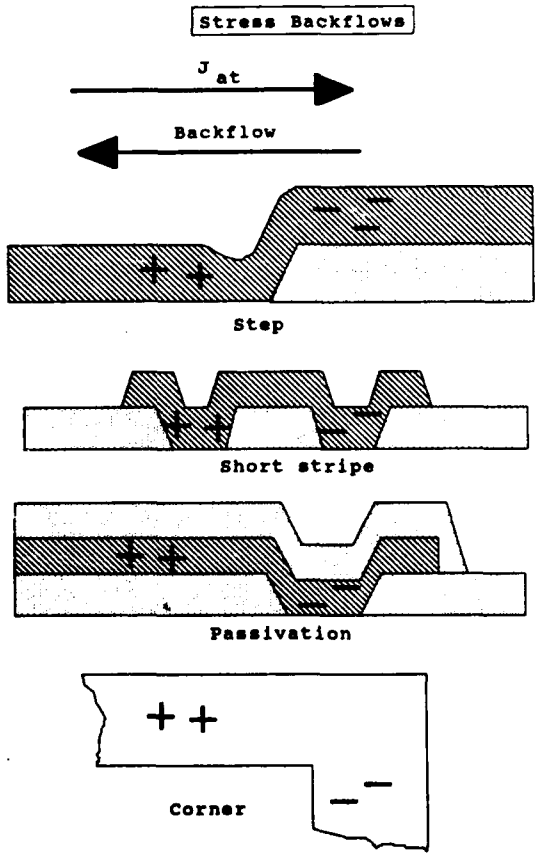


Figure 5. Beneficial effects of backflows in various common geometries in VLSI. Note that these effects become appreciable at low current densities.

LIFETIME MEASUREMENTS

Lifetime measurements follow traditionally the pioneering work of Black(20). A sample of test conductors of specific thickness, width and length are tested isothermally under constant current conditions and their failure time (when open) recorded. The results are plotted on a Log-Normal scale, such as Figure 6, and the parameters reported are time to 50% failures (MTF) and the standard deviation, . By the nature of such curves, the actual instantaneous failure rate is not constant and has to be calculates for each time interval(21). Typical lifetime measurements are done at 175-250°C at current densities between $8x10^5$ and $3x10^6$ A/cm^2. Following Black(16) the lifetime at normal operating temperatures and current densities is calculated using an empirical equation,

$$MTF = \text{const} \times j^{-n} \times \exp(Q/kT) \qquad (7)$$

where,

MTF - Mean time to failure.
const - A constant.
n - An empirical constant
Q - Activation energy

The value of n as well as Q was a subject of numerous measurements(7,22). Reasonable values to use for non-alloyed aluminum are n=1.7-2 and Q=0.5-0.55eV. Figure 6 shows a typical Log-Normal curve for electromigration lifetime measurements.

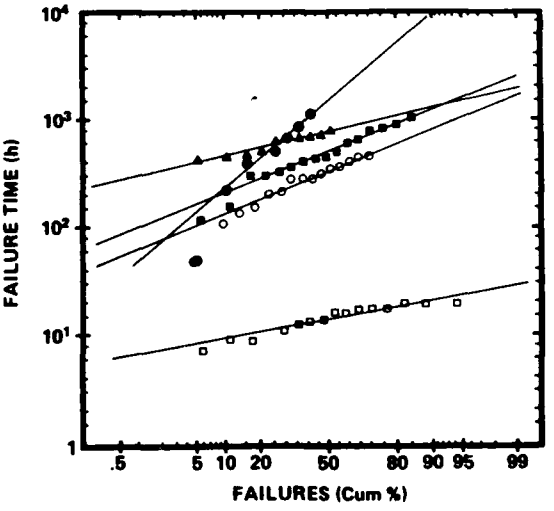


Figure 6. A typical Log-normal curve depicting the cumulative failures vs. time. This example shows the variations in lifetime of Al-Cu strips formed by various depositions(27).

Based on lifetest results, some specification allow current densities of $2x10^5$A/cm^2 for unpassivated pure aluminum, or aluminum alloys, while others limit the current density to $1x10^5$A/cm^2.

LIFETIME SIMULATIONS

Lifetime predictions have been a subject of several papers (18,21,23,24). The subject is quite complicated not only because of its statistical nature but also because of the many parameters that affect the lifetime. The lifetime is affected by a variety of parameters such as the current density, material, composition, microstructure, overcoats, length, width, thickness, surface diffusion, temperature, temperature gradients and even ambient(25). Most attempts were one dimensional in nature(21,23,24), but recently, there have been some attempts to address some of these aspects by constructing a 2 dimensional conductor microstructure and simulating the effects of current passage. These attempts are just a first step to the solution of a very complicated problem. The backflow effects are also complicating the lifetime predictions and have yet to be included in the 2 dimensional simulations. As mentioned before, the backflow effects become important at the current densities which are actually used during device operation but are practically absent at the high current densities used during accelerated life tests.

Finally, the practical problem of the relation between actual circuit lifetime and test pattern lifetimes has never been satisfactorily resolved.

REDUCTION OF ELECTROMIGRATION

The drive toward electromigration reduction has been the subject of several investigations since the discovery of electromigration failures in thin films. The methods proposed to the electromigration problem are several:

1. Additions of alloying elements.
2. Passivation layers
3. Sandwich structures of a refractory metal and the conductor.
4. Use of refractories.

The first method for electromigration reduction was discovered by IBM researchers (26) who found that additions of copper to aluminum increased its lifetime by about two orders of magnitude. A recent careful study of alloying effects was published recently(27) showing that the lifetime of alloy films is dependent on their deposition conditions.

Shortly after the discovery of alloying effects it was suggested that passivation layers are also reducing electromigration damage. The use of a restraining passivation layer should be especially beneficial in fine line conductors where the restraining effect are causing higher backflows and reduced diffusivity.

The use of refractory sandwich, underlayers and overlayers was proposed by Howard et al(28). He showed a many fold increase in lifetime for various film combinations. A revival of his work was reported recently(29,30).

Finally, the use of refractories to replace aluminum gained some momentum with the intro-duction of some products using that scheme by both Hewlett-Packard and General Electric(31).

CONCLUSIONS

The subject of electromigration failure, testing and simulations is quite complex. There is a large scatter and disagreement in test results. The reasons for this situation is the complex nature of the events leading to failure. It is our hope that in the future we will be able to model more closely the electromigration failures and at the same time use more electromigration resistant conductors. Electromigration failures do not comprise a large percentage of the field failures of today primarily due to the adherence to safe design rules.

REFERENCES

1. I.A. Blech and H. Sello in "*Physics of Failure*", **Vol 5**, T.S. Shilliday, J. Vaccaro eds., RADC (1966) 466.
2. D. Chhabra and N. Ainslee, *IBM Technical Report* **No.22**, 419 (1967)
3. H.B. Huntington and A.R. Grone, *J. Phys. Chem Solids* **20**, 76 (1961).
4. A.R. Grone, *J. Phys. Chem Solids* **25**, 335 (1964).
5. I.A. Blech, *J. Appl. Phys*, **47**, 1203 (1976).
6. F.M. D'Heurle and P.S. Ho in "*Thin Films - Interdiffusion and Reactions*", J.M. Poate, K. N. Tu, and J.W. Mayer eds., John Wiley, NY (1978).
7. H.U.Shreiber and D. Grabe, *Solid State Elect.*, **24**, 1135 (1981).
8. J.R. Lloyd, *21st International reliability Physics Symposium*, (1983) pg 208.
9. S. Vaidya and A.K. Sinha, *20th International reliability Physics Symposium*, 1982 pg 50.
10. P.A. Gargini, C. Tseng and H.M. Woods, *20th International reliability Physics Symposium*, (1982) pg 66.
11. G.J. Chern, W.G. Oldham and N. Cheung, *IEEE Trans. elctr. dev.*, **ED33**, 1256 (1986).
12. S.D. Steenwyk and E.F. Kankowski, *24th International reliability Physics Symposium*, (1986) pg 30.
13. Y.E. Strausser, B.L. Euzent, R.C. Smith, B.M. Tracy and K. Wu, *25th International reliability Physics Symposium*, (1987) pg 140.
14. I.A. Blech, *J. Appl. Phys.*, **48**, 473 (1977).
15. I.A. Blech and C. Herring, *Appl. Phys. Let.*, **29**, 131 (1976).
16. I.A. Blech and K.L. Tai, *Appl. Phys. Let.*, **30**, 387 (1977).

17. E. Kinsbron, *Appl. Phys. Lett.*, **36**, 968 (1980).

18. K. Nikawa, *19th International reliability Physics Symposium*, (1981) pg 145.

19. C.A. Ross and J.E. Evetts, *Scripta Met.*, **21**, 1077 (1987).

20. J.R. Black, *IEEE Trans electron. devices*, **ED-16** 338 (1969).

21. P.B. Ghate, "Thin film electromigration", tutorial given at *20th International reliability Physics Symposium*, (1982).

22. K.A. Danso and L. Tullos, *Microelectronics & Reliab.*, **21**, 513 (1981).

23. J.D. Venables and R.G. Lye, *10th International reliability Physics Symposium*, (1972), pg 159.

24. R.A. Sigsbee, J. *11th International reliability Physics Symposium*, (1973), pg 301.

25. D.-Y. Shih and P.J. Ficalora, *16th International reliability Physics Symposium*, (1978), pg 268.

26. I. Ames, F. D'Heurle and R. Horstmann, *IBM J. Res. dev.*, **14**, 461 (1970).

27. P. Marchant and T. Cass, *22th International reliability Physics Symposium*, (1984), pg 259.

28. J.K. Howard, R.F. Level and P.J. Smith, *J. vac. Sci. and Techno.*, **13**, 68 (1976).

29. J.M. Towner, A.G. Dirks and T. Tien, *24th International reliability Physics Symposium*, (1986), pg. 7.

30. J.A. Maiz and B. Sabi, *25th International reliability Physics Symposium*, (1987) pg 145.

31. M.J. Kim, D.W. Skelly and D.M. Brown, *25th International reliability Physics Symposium*, (1987), pg 126.

AN OVERVIEW OF ELECTRONIC PACKAGING

Donald P. Seraphim, Donald E. Barr, John M. Brauer
International Business Machines Corporation
Endicott, New York, USA

Packaging is the science and engineering of interconnecting (wiring) for electrical functions. An electronic package provides a suitable environment for signal distribution in addition to the electrical interconnections. The environment must provide proper power and distribution, effectively remove heat generated by the electrical components, and supply the required mechanical integrity and protection from adverse environmental conditions.

The components that are interconnected in a package, semiconductor devices, resistors, inductors, capacitors etc comprise only about 25% of the system cost while the substrates, printed circuits, power supply, support interconnection and cooling make up the remaining 75%. Integration of the components has the favorable impact of decreasing the number of substrates, printed circuits, the support interconnections, and provides a corresponding reduction in cabinets and frames. Therefore, packaging trends tend to follow the semiconductor integration trends, factors of 3 or 4 increase in integration at the semiconductor level provide the initiative for an equal increase in packaging density and a commensurate lowering of costs for a system function. In addition, interconnection costs at the semiconductor level are in the range of fractions of a cent while those at the first package level are in the range of 1 to 10 cents per wire and cabling costs are an order of magnitude higher. Therefor there is a tendency to drive interconnections onto the chip wherever possible to the extent that chip yields will permit the integration.

The electronic packaging industry may be segmented into consumer products, small systems, mid-range systems and performance oriented large systems and super computers.

The industry segments have substantially different function and component characteristics. Therefor, the factors that bring about the characteristics and trends in packaging are discussed here.

CONSUMER PRODUCTS (calculators, controllers for cameras, compact discs, toys).

- Typically low integration level components.
- Small to large numbers of components per system.
- Often space constrained packaging.
- High volume manufacturing.

This part of the industry, Figure 1, has been growing very rapidly with numerous benefits derived from Technology trickling down from the more elaborate systems technology. In particular, the heavy use of small microprocessors, memory

chips, and a variety of interconnection technologies has allowed the rapid introduction of packaged electronics to the consumer. In fact, the original process practiced in the 1960's for forming printed circuit interconnections by screening resist patterns on copper panels is a predominant method of forming interconnections in this market segment. On the other hand, this segment has been an entry point for tremendous statistical learning for a variety of components; particularly keyboards, liquid crystal displays and voice components. The volumes are very large and provide manufacturing experience as an added benefit.

LOW END SYSTEMS (personal computers, work stations, printers

- Microprocessor driven
- Highly competitive rapid design iterations
- Mostly off the shelf components with wide variety
- Few highly integrated chips
- Quality products at high volume and low cost
- Printed circuit interconnection 2 to 4 signal layers

Low end systems predominately use microprocessors with 8, 16 and 32 bit architecture. There have been three generations of integration of the logic components in just ten years, while the memory doubles approximately every two years. Most components have relatively few contacts (commonly called input/output or I/O), and therefore, require a low density of interconnection. This allows the few processor and adapter chips, usually 4 to 10 highly integrated components, to be spaced far enough apart on the printed circuit board to permit the required interconnection circuit wiring.

The rapid growth in performance of these systems, however, drives the lead counts (I/O) upwards while maintaining low interconnection costs. Thus dual in-line component configurations have been replaced by components with leads on four sides with decreased spacing which reduces the printed circuit board area needed. To maintain simplicity in the circuit board area, the holes are made much smaller to allow more space for circuit wiring. Instead of inserting the component leads in the holes on the card, they are connected by soldering to a surface pad. All of these miniaturizing techniques improve the electronic performance of the system but the main initiative is to lower the cost as a function of electrical performance. Since these devices are surface mounted they are also somewhat easier to automatically handle than pinned devices. Increased integration levels are more important in driving costs down by elimination of components and decreasing the number and complexity of the assembly operations. For example, a few very highly integrated pin grid array devices may be used in these systems with connector blocks integration into the third dimension without use of expensive structural frames, Figure 2 and 3.

MID RANGE SYSTEMS (Disc systems, processors)

- High levels of integration
- Less component variation
- Interconnectivity
- Multiple busses
- Incremental features

These systems usually contain substantial memory and are also typified by performance through large logic depth at the card level. They achieve their incremental performance levels by three dimensional packaging of cards plugging into printed circuit back panels containing connector blocks.

Historically these systems provided even the highest performance systems by the use of a large number of card-on-board frames cabled together. This trend has continued with the use of very large electronic function cards and with connectors attached to both ends of the cards to achieve the function growth.

The increase of I/O's on both the integrated circuit components and at the card level is RENT like, Figure 4, that is, the I/O's increase exponentially with the circuit counts, C.

$$I/O = A\,C^{1/2}$$

The pre-exponential constant factor A ranges from 2 to 4.

The high integration levels in mid range systems require dense I/O configurations at the component level of package. Space is often at a premium. These area array components are typically multi-layer ceramic pin grid arrays PGA or the simpler metallized ceramic pin grid arrays. There is some use of high density surface mount devices, PLCC, and ceramic leaded chip carriers in mid range systems. The high circuit count at the card level drives the use of complex performance oriented connector systems that need tight tolerance control with a complex frame system. The growth of integration at the card level may require connection densities beyond that which can be achieved on one end of a card -- thus zero insertion force systems (ZIF) have been developed with contacts to back panels at both ends, Figure 5 and 6.

The printed circuit density increase is achieved by increasing circuit wiring density within each layer and/or increase the number of layers. The circuit density increase is in direct proportion to the number of I/O per component and the spacing of the components -- thus the printed circuit complexity continues to increase as integration proceeds at the chip level. One might think that as more interconnection occurs on the chips, less would be needed on the cards. That would be true if the circuit count at the card level remained constant. But the integration level continues to grow at all levels with increasing system size and performance dramatically increasing year after year,

HIGH END SYSTEMS (Computers, Super-computers)

- Performance oriented circuits/ reachable in a nanosecond
- Large functional units
- High power density
- Special cooling configurations
- Performance cabling systems

The drive for logic depth, the number of circuits which can interact in a unit time, i.e. circuits per nanosecond -- places extreme demands on technology. The chips are advanced to the highest levels of integration possible along with power driven performance. Then they are placed as close as possible in large arrays -- multilayer multichip modules with flip chip connections are typical of the densest configurations. Miniature ceramic leadless chip carriers housed on multilayer ceramic substrates with thin film interconnections also provide comparable density. These are followed by surface mounted chip carriers on printed circuit cards. In this latter case, both sides of the card may be used for assembly of devices and the cards may be much larger than the multilayer ceramic carriers. Obviously there are a large variety of configurations which will provide adequate logic depth.
The main mechanical feature of the ceramic substrate is its close coefficient of thermal expansion CTE match to the chip allowing short interconnections at the interface between them. The low inductance and impedance match of that interconnection is important for the performance at this interface. The interconnection capabilities of the multilayer ceramic, small vias and many layers favor spacing the chips very close together -- thus, the power densities become extreme and special cooling configurations are required, Figure 7.

On the other hand, the significant features of the organic multilayer technology are dielectric constant material and the power delivery capability provided by thick layers of copper. The mismatch in CTE between the chip and the organic panel creates the need for substantial strain relief in the interfacing interconnection leads -- thus, the electrical performance, although better in the interconnection layers of the printed circuit cards, than in the module, is degraded by the impedance mismatch and high inductance of the interface I/O.

The cooling configurations are unique for each of the variety of structures. Large heat sinks or combinations of conduction cooling and liquid cooling are required. Since very large circuit functional units are needed containing many

chips, very large numbers of I/O from the card or the multichip module to the next level package are required. Assuming, for example, several thousand I/O as shown earlier with the RENT relationship are required, then grid arrays become appropriate. The alternative, very dense edge arrays on cards require many inches of card edge and the tolerances are difficult to achieve. Thus very rigid precision frames must accompany these packages, Figure 8.

SUMMARY OF TRENDS

By keeping with the above architectural, performance and cost considerations in mind, future trends are relatively straight forward.

In the consumer applications, the tendency is to stay in plastics; plastic organic chip carriers PLCC or TAB, wire bonded chips directly on a simple card. The cards will remain the simplest possible with single layers of interconnections with mass production screening processes or be replaced with molded simple card configurations. Low End systems will gradually benefit from some of the technologies established for the high end. As time permits integration will remove the large numbers of components simplifying assembly and allowing the full advantage of integration to prevail. The very large increase in functional capability will require use of high density printed circuit interconnections. Competitiveness will establish a trend for more performance which will be met without impacting cost as the technology learning curve matures for increasing densities.

Mid Range Systems will continue to grow in performance with increasing logic depth at the card level. Thus interconnection demands will continue to increase. This may be handled by increasing use of multi-chip modules along with denser grid arrays and TAB peripheral surface mount components, a trend which is already set in place. The increased function at the card level will continue to press the use of larger numbers of card I/O. So far, this can only be achieved by using large cards with more perimeter use for I/O. Since the large circuit densities will be supplied by high integration levels of CMOS rather than bipolar circuits, cooling will continue to be satisfied by conventional means, heat sink enhancements and high volumes of air. The performance of the interconnection will tend to be more stringent in its tolerances. The impact of electronic coupling between interconnection lines at high density will call for lower dielectric constant materials.

High End Systems have been growing in size by an order of magnitude in less than 10 years approximately at the same rate of integration occurring in the chips. Although the power per circuit has continued to decrease, the integration increase is more rapid. Thus, chip power and the methods of cooling are limiting factors in tightly packaging many bipolar chips. Further innovations in liquid cooling will be necessary to remove the heat generated by the chips. The limitations of the ceramic and printed circuit board technologies in signal performance will dominate development efforts. Interconnection densities must increase in proportion to the chip I/O -- approximately a factor of 3 for 9X increase in integration. Since interconnection resistance is an important factor in these larger systems interest will increase in thin films and in liquid cooling in nitrogen, which decreases resistance in copper by a 5X factor. In addition, the advantages of very high mobility of the carriers in silicon CMOS circuits and in GaAs will complement the improved packaging performance attainable.

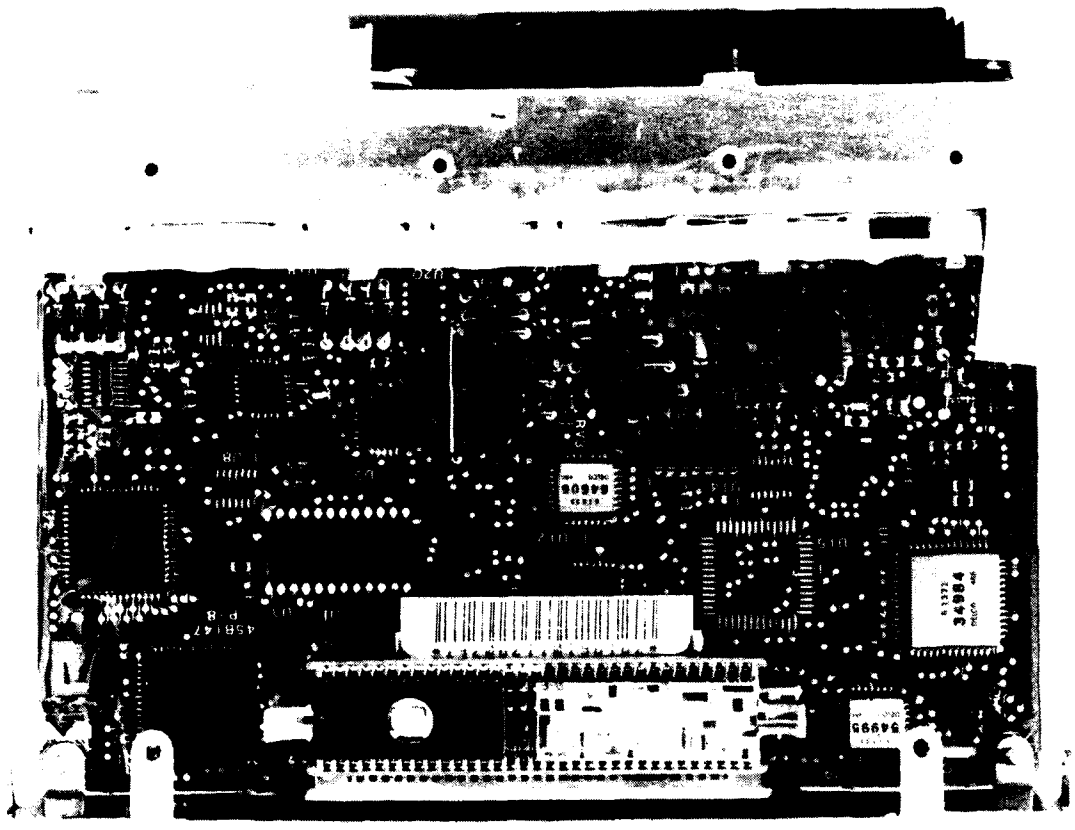
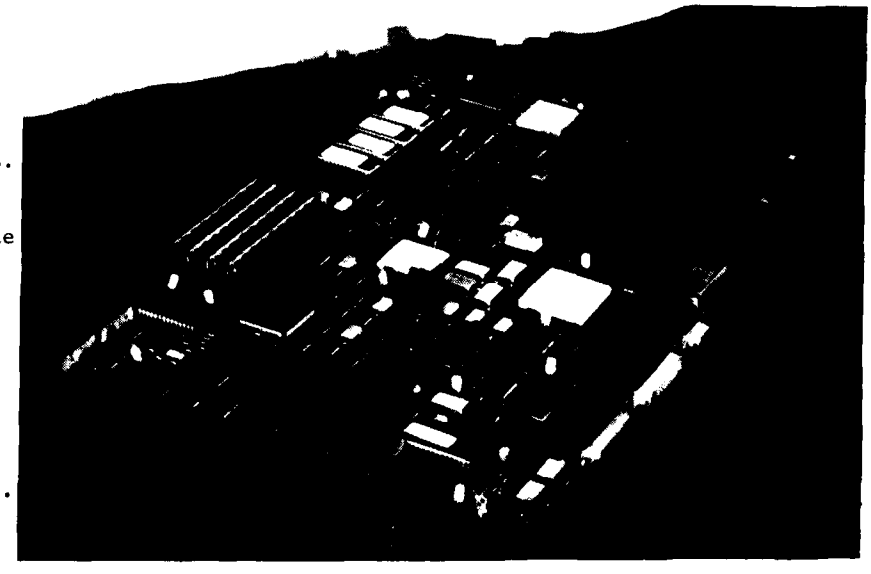


Fig. 1 - Automobile Controller unit containing several simple microprocessors and many non-active components. (Courtesy of General Motors Delco)

Fig. 2 - A Printed Circuit Board from a PS-2 Personal Computer, courtesy of IBM demonstrating the wide variety of components used in low-end systems. Shown here are Dual-In-Line (DIP), Small outline J leaded surface mount components, Flat Packs, Plastic Leaded Chip Carriers PLCC, and connector blocks for connection in the third dimension. This printed circuit board had four layers of wiring and two ground and power layers.



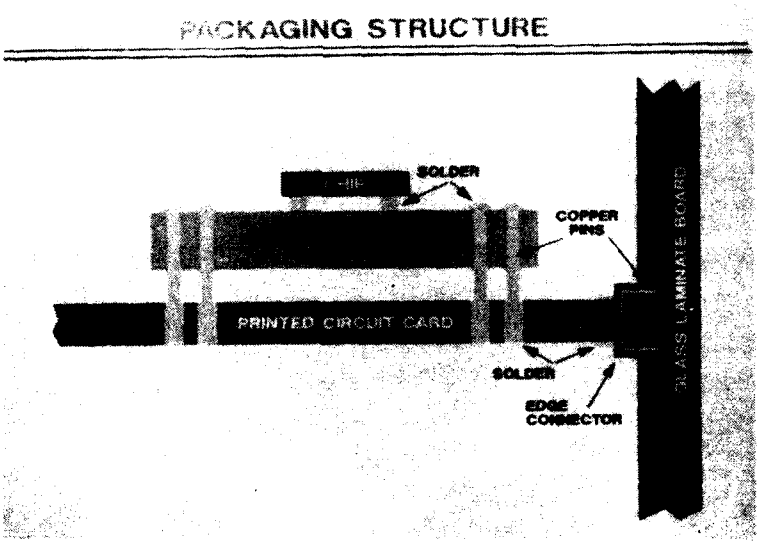


Fig. 3 - A cross section showing hierarchy of interconnection from chip to printed circuit back panel. The chip joint interface to the first level package here is a solder ball or "flip chip" with chip face down. A common practice in the industry is for chips face up with wire bonding joining the chip terminals to lead frames or ceramic substrates. The first level package is soldered pin in hole here and in other cases is surface mounted to the printed curcuit card. The card is plugged into a connector contact block in a circuit back panel.

Fig. 4 - A compilation of data on chips, modules, cards and boards relating I/O to the number of circuits contained.

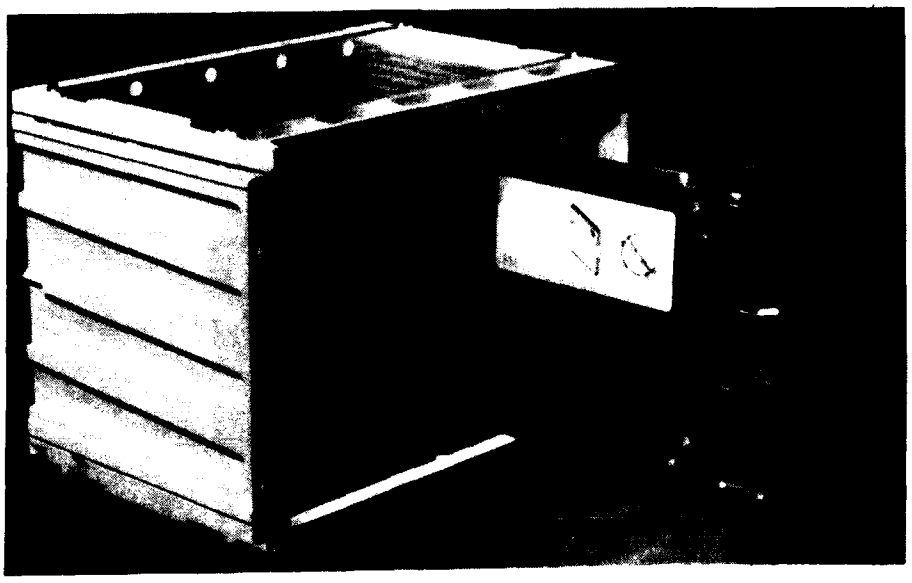


Fig. 5 - A printed circuit card on printed circuit board packaging hierarchy. A rigid frame system (in contrast to Fig. 4) holds the printed circuit back panels in precise position for terminal pads on the card to mate with contacts supported in the printed circuit board. The cards with terminals on both ends slide into place in the contact system with zero insertion force (ZIF). The cards typically have 4 wiring planes and as many as 4 power planes while the printed circuit boards have as many or more layers. Several of these frames may be cabled together in the larger systems.

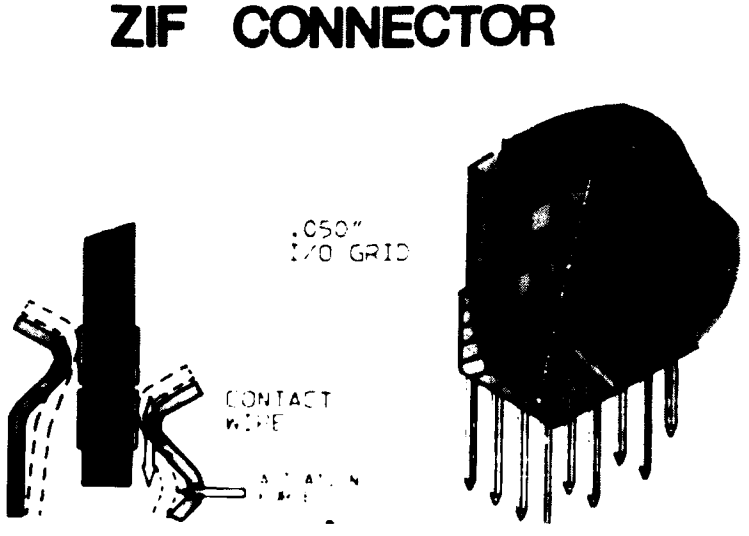


Fig. 6 - A zero insertion force connector contact. The high reliability of the ZIF contact systems is established through careful parametric studies which include contact shape, forces, wipe, metallurgy and the effects of extremely hostile environments.

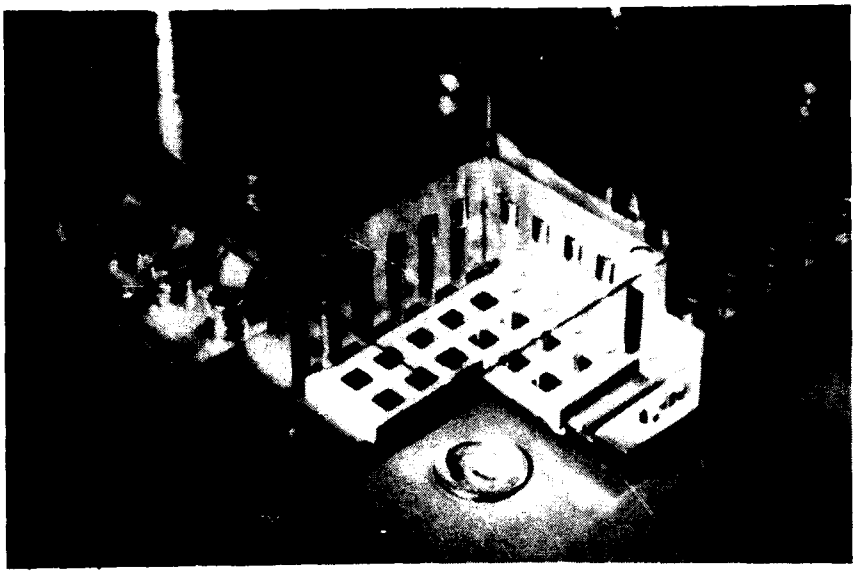


Fig. 7 - A closed packed array of chips on a multilayer ceramic carrier with its cooling system. The high density of chips necessitates a liquid cooled environment to remove about 500 watts of power. The interconnection and voltage distribution in the module is provided in 30 layers of interconnetion. The interface between the moduie and printed circuit board is an area array of connector pins which plug into a bifurcated zero insertion force connector system in the printed circuit board.

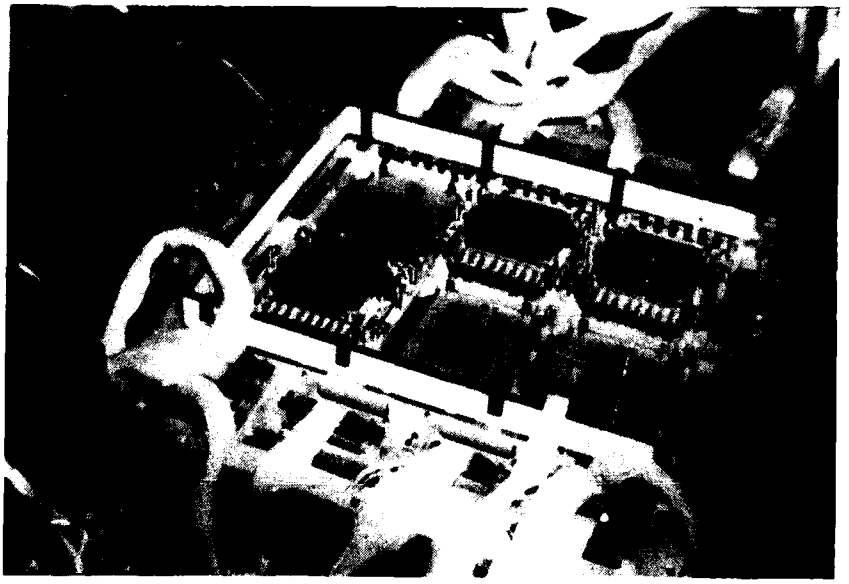


Fig. 8 - A high end system board which interconnects 6 thermally cooled modules (TCM). The number of chips on this board is in the range of a hundred. Approximately a mile of 100 μm wide interconnection is provided in 8 layers. Power distribution is provided in 12 additional layers. The combination of layers in the TCM and the printed circuit board is greater that 50. Printed circuit boards which have comparable function comprised of single chip modules rather than TCMs also require 40 to 50 layers.

BIBLIOGRAPHY

Microelectronics Packaging Handbook, VanNostrand-Reinhold (1988), R.R. Tummala and E.J. Rymaszewski, Editors.

Principles of Electronic Packaging - McGraw Hill, N.Y. (1989), D.P. Seraphim, R. Lasky, C.Y. Li, Editors.

Advanced Thermoset Composites - VanNostrand-Reinhold Co., N.Y. (1987), J.M. Margolis, Editor.

Handbook of Microelectronic Packaging and Interconnection Technologies, Electro-chemical Publications, Ayr. Scotland (1985), F.N. Sinnadurai, Editor.

HIGH SPEED, HIGH DENSITY, LOW LOSS INTERCONNECTIONS FOR VLSI, VHSIC, AND GaAs ON A VARIETYOF SUBSTRATES

K. Jayaraj, T. J. Moravec, R. Jensen, F. J. Belcourt, R. Sainati
Honeywell Sensors and Signal Processing Laboratory
Bloomington, Minnesota 55420, USA

ABSTRACT

There has been a large increase in the efforts by many people over the last five years to develop thin film techniques and use them, possibly in combination with plating and other Printed Wiring Board (PWB) techniques, to fabricate multichip packages (MCP) or boards for advanced digital ICs, i.e., VLSI, VHSIC, and GaAs. Honeywell pioneered the development fine line photolilthographically defined thin film copper conductors and polyimide dielectric (Cu/IP) interconnects on ceramic substrates. Although extension of IC techniques on silicon is also being developed for similar packaging, the Cu/PI technology is becoming the most popular as evidenced by the packaging in the NEC supercomputers. We review MCP technology, present some of its features and physical properties, and discuss the requirements for substrate materials, and the effects of closely spaced conductor lines on electrical performance, especially at high speeds (1 GHz). We also outline the effect of high fanout density (on a busline) and power/ground noise induced by a large number of simultaneously switching I/O's.

INTRODUCTION

New approaches are needed and are being developed to interconnect the latest generation of VHSIC and GaAs ICs. High-density interconnects are necessary to interconnect a large number of high I/O devices such as VHSIC ICs. To interconnect high-speed GaAs devices, impedance controlled interconnects designed as transmission lines (with proper considerations for propagation delay, crosstalk, attenuation and rise time degradation) with proper termination are required.

Conventional single chip packages (SCP) pose severe limitations on system speed due to long interconnects needed between chips, and also on the maximum attainable chip density. To overcome the limitations of conventional packaging approaches, several companies are developing multichip packaging approaches (1-7). Compared to SCP on printed wiring boards, the MCP approach permits greater chip density (since bare die are used), reduced power consumption and propagation delay (due to short interconnects), and higher system reliability (due to fewer external connections involved).

The high interconnect density of the MCP approach requires the ability to define high resolution conductor lines and dielectric vias in multiple layers on large substrates. Interconnects must be narrow, high aspect ratio lines of a highly conducting material to maximize interconnect density and to minimize resistive losses. Dielectric layers must be thick to minimize interconnect capacitance, and the material must have a low dielectric constant to minimize propagation delay, interconnect capacitance and crosstalk between signal lines.

Finally, improved cooling techniques are required to handle increased power density due to increased chip density.

Honeywell and other companies have developed thin film multilayer (TFML) technologies to satisfy the above requirements. Honeywell's approach is based on copper conductor and polyimide dielectric materials. Polyimide is an ideal material as a dielectric for MCP due to its low dielectric constant (3.2, depends on relative humidity), and the ease of application. Copper can be deposited by a variety of techniques, and is ideal because of its high conductivity. Polyimide is applied by either spinning or spraying, and patterned using reactive ion etching with SiO_2 as the masking material. Copper is sputtered and patterned by wet etching techniques. Conductors (5 μm thick and 25 μm wide) are designed as offset striplines and are spaced on a 100 - 125 μm pitch. The dielectric layers are typically 25 μm thick. The details of Honeywell's approach are presented elsewhere (1-4).

There are a number of issues that must be addressed to implement a successful TFML approach. An attempt is made here to address the problems caused by large substrates and the effects of narrow, closely-spaced conductor lines on electrical performance. We also briefly outline the effect high fanout density on the bus line and the effect of large number of simultaneously switching I/O's on induced power and ground noise.

SUBSTRATE CONSIDERATIONS

SUBSTRATE MATERIAL: One of the advantages of the sequential TFML process approach is that it can be implemented on a wide variety of substrates, including ceramics (multilayer co-fired or tape cast), metals, and silicon wafers. The choice of substrate will be dictated by the overall package design and system requirements. For example, multilayer ceramic substrates can provide additional power and ground planes and can have vias through the substrate to a pin or pad grid array on the bottom, whereas metal or silicon substrates are generally mounted in a secondary package (such as a flatpack), with peripheral I/O connections to the top surface of the substrate.

The ideal substrate material will have some of the following characteristics: It should have a high thermal conductivity for good heat dissipation from the ICs. It should have a high modulus of elasticity and a coefficient of thermal expansion (CTE) close to polyimide (typically 20 - 40 ppm/^{o}C) to minimize warpage due to thermal expansion mismatch with polyimide (discussed later). However, a CTE close to silicon (2.3 ppm/^{o}C) may be desirable for flip chip bonding or the attachment of large die. The material should be chemically inert to the wet etchants and plasma processes used in fabrication. The surface should be smooth, flat, and defect-free, which usually requires grinding or polishing processes. Finally, the ideal substrate should be light weight, inexpensive, and available or machineable to a variety of shapes and sizes.

Ceramics (particularly alumina) have been widely used as interconnect substrates because they have good mechanical strength and stiffness, they are chemically inert, they can contain multilayer metal patterns, and they are commonly used in hermetic, high-reliability packages. More recently-developed ceramics such as SiC and AlN also have high thermal conductivity. A wide variety of metal substrates meet several of the requirements discussed above. Metal substrates investigated at Honeywell, and their advantages, include aluminum (light weight, low cost, machineable, high thermal conductivity), copper (high conductivity, low cost), molybdenum or tungsten (high modulus, low CTE), Cu-clad Mo, and steel (low cost, machineable). Finally, silicon is becoming a popular substrate because it has a relatively high thermal conductivity and a perfect CTE match to silicon die, it has a high-quality surface finish, and it is widely available in standardized sizes that are adaptable to IC process equipment. However, silicon has a low modulus and a poor CTE match to polyimide, necessitating very thick substrates to prevent excessive warpage.

SUBSTRATE AREA: There is a continuing demand for larger area substrates in order to capture more functionality in a multichip package. The practical limit to substrate size is often dictated by equipment limitations. Current IC process equipment is limited to 6-inch or sometimes 8-inch wafers; this limits square interconnect substrates to 4 - 5.5 inches. As thin film equipment for larger-area devices such as flat panel displays becomes available, this equipment may be applied to interconnect substrates.

A more fundamental limitation to substrate area and thickness is imposed by the thermal expansion mismatch between polyimide and the substrate, which causes tensile stress in the polyimide film as it cools from the cure temperature, resulting in warpage of the substrate. This warpage

or camber complicates further patterning processes (particularly photolithographic alignment) as well as die attach and substrate assembly processes.

It can be shown that the camber of a round substrate is related to the geometries and film properties according to the following expression:

$$C = D^2/4\ (3\ t_f/t_s^2)\ (E_f/E_s)\ (1-\nu_s/1-\nu_f)\ (a_f - a_s)\ (T_c - T_r)$$

where D is the substrate diameter, t_f and t_s are the thickness of the polyimide film and substrate, respectively, E_f and E_s are the Young's modulus of the film and substrate, ν_f and ν_s are Poisson's ratios, a_f and a_s are CTEs, and T_c and T_r are the polyimide cure temperature and room temperature, respectively. The camber increases linearly with substrate area and polyimide film thickness, and is worse for low modulus substrates with a low CTE. Silicon is such a material; for standard 4" diameter x 0.020"-thick silicon wafers, a typical multilayer polyimide thickness of 4 mil would cause an unacceptable camber of 45 mil. The silicon substrate should be 60 mil thick to produce a more acceptable camber of 5 mil. In applications where there is a limitation to the overall thickness of the substrate and interconnects, a high modulus substrate (e.g. Mo, W) has a significant advantage.

The second fundamental limitation to substrate area is yield. To a first approximation, the yield for defect-size dependent faults (such as conductor line opens and shorts) depends exponentially on the substrate size and defect density according to the relationship:

$$Y = \exp(-k\ (L/x)^2)$$

where L is the side length of the substrate and x is the minimum critical defect size, which is related to the minimum feature size of the conductor pattern. Since the yield depends exponentially on the ratio $(L/x)^2$, the yield for pattern in 10 μm lines on a 10 cm substrate is equivalent to patterning 1 μm lines on a 1 cm chip, which approaches the limits of current IC production technology. Because of this severe yield constraint, in-process testing methods such as high speed capacitance probing or voltage-contrast scanning electron microscopy, and repair techniques such as laser etching or writing, may be required to obtain acceptable manufacturing yields.

ELECTRICAL CONSIDERATIONS

ATTENTUATION AND RISETIME DEGRADATION: Selection of an appropriate interconnect technology involves consideration of many factors, one of which is electrical performance. The smaller line geometries of today's thin film interconnects result in increased line densities and decreased chip spacings. A disadvantage is the possibility of risetime slowdowns and decreased signal levels because of higher line loss plus increased crosstalk due to closer line spacings. We investigated the electrical performance of several thin film interconnect technologies through simulation to determine waveform fidelity and crosstalk. The TFML technologies of AT&T, Raychem, and Honeywell were studied. Published dimensions and material parameters (10 μm X 2 μm lines with 20 μm pitch for AT&T, and 25 μm X 6 μm aluminum conductors on 75 μm pitch with 10 μm dielectric layers for Raychem) were used to determine line electrical characteristics such as impedance and loss. Simulations were performed for clock signals with risetimes of 10, 1, and 0.1 ns (representing 10, 100, and 1000 MHz clocks) propagated down 3, 10, and 20 cm long lines. In all cases, the lines were assumed to be terminated in their characteristic impedance.

Simulations for the 10 MHz clock show no significant signal degradations until line lengths approach 20 cm. Figure 1 shows the results which indicate a significant risetime slowdown and signal reduction for the very fine AT&T line geometry. The other two technologies show much less signal degradation. For the 100 MHz clock, the AT&T interconnects show (Figure 2A) serious degradations at 10 cm, whereas the others (although not shown, the Raychem results are similar to the Honeywell waveforms) begin to show degradation at 20 cm line lengths, (see Figure 2B). Results for the 1000 MHz clock, Figure 3, again show signal degradation for the AT&T geometry and also indicate the usability of the Honeywell technology to at least 10 cm (but not 20 cm). These results are not surprising since both risetime slowdown and signal level reduction are caused by increased line resistance. Narrower lines will have unavoidably higher resistance, but offer a higher interconnect density. Thus some care must be taken when selecting an interconnect technology if either fast risetimes or long line lengths are required.

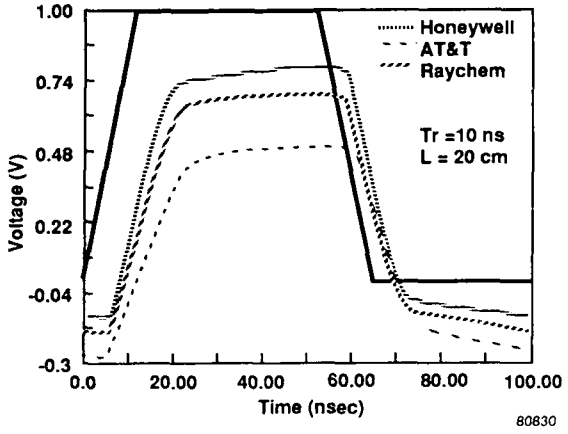


FIGURE 1 Waveforms for a 10 MHz clock Propagated through a 20 cm long line

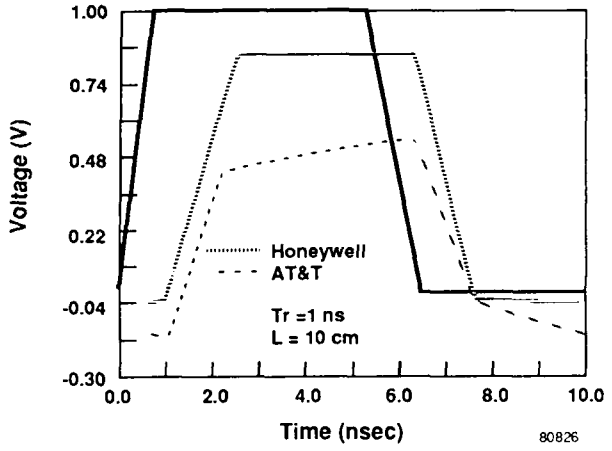


FIGURE 2A Waveforms for 100 MHz clock propagated through a 10 cm long line

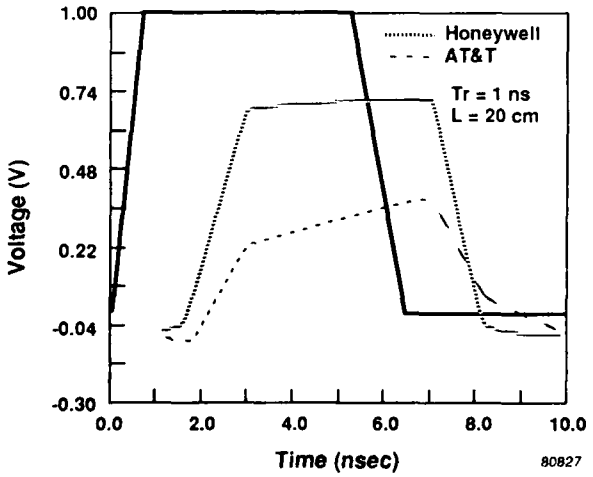


FIGURE 2B Waveforms for 100 MHz clock propagated through a 20 cm long line

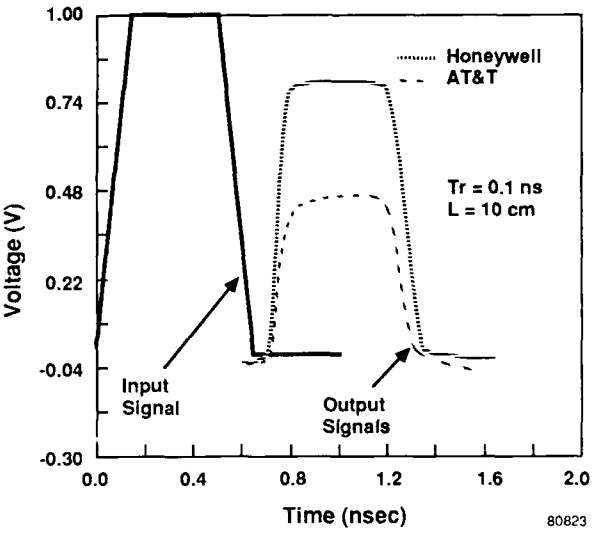


FIGURE 3 Waveforms for a 1000 MHz clock propagated through a 10cm long line

CROSSTALK: Additional simulations were performed to determine crosstalk levels using representative line spacings. Table 1 shows peak backward and forward crosstalk levels for two 10 cm long parallel lines of each technology as a function of risetime. For reference, the input is assumed to have a 0 to 1 volt swing. As expected, closer line spacings result in higher crosstalk levels. The effect of crosstalk values such as 115 mV for some of the cases is best determined by system considerations. Still, the possibility for higher crosstalk needs to be traded-off with line density and line width needs. The very low crosstalk levels for the Honeywell technology result, in part, from the use of a stripline structure for the interconnects. The other two approaches are essentially buried microstrips. The addition of a second ground plane in Honeywell's technology greatly reduces crosstalk.

THE BUS ISSUE: Electronic systems usually consist of several distinct components such as memories, processors, and peripherals that must share and exchange data and instructions. This is accomplished using a physical medium called a bus. A bus line interconnects many points in a system with a common signal path. Generally, drivers and receivers are located at each end of a the line, with additional receivers and drivers connected to the bus at discrete intervals along the line. System busses normally utilize a single-ended protocol, as versus a differential scheme which consists of a two-wire driving system for

minimization of noise. However, the differential approach presents problems regarding driving compatibility and the introduction of crosstalk noise due to the data switching of only one line.

Table 1

Peak Crosstalk Values for 10 cm long lines of each Technology as a function of Signal Risetime

Signal Risetime (ns)	Technology	Peak Crosstalk Value: Backward Crosstalk (mv)	Peak Crosstalk Value: Forward Crosstalk (mv)
10	Honeywell	0.87	0.013
10	AT&T	59	4.6
10	Raychem	10	1.4
1	Honeywell	1.2	0.018
1	AT&T	115	9.9
1	Raychem	19.9	3.5
0.1	Honeywell	1.2	0.018
0.1	AT&T	115	71.3
0.1	Raychem	19.9	34.7

line lengh = 10 cm
0 - 1 volt input signal

High-speed, single-ended distributed busses usually have a high fanout density which introduces additional capacitance to the intrinsic line capacitance. When this happens, electrical parameters such as characteristic impedance, propagation speed, and noise immunity are impaired. The characteristic impedance of a distributed bus line is modified in the following manner;

$$Z_0' = Z_0/\mathrm{sqrt}(1 + C_d/C_0\ell)$$

where Z_0 is the unloaded line impedance, C_d the distributed load capacitance, C_0 the intrinsic line capacitance, and ℓ the length of the line. Fanout loading along the line and the intrinsic line capacitance due to length tends to change the characteristic impedance. The fine-line copper/polyimide package approach minimizes the intrinsic capacitance by maintaining short interconnect paths between chips within a small wiring area. The loading on a bus line should be evenly distributed; this provides a constant impedance along the line which minimizes reflections. Copper/polyimide, with its shorter lines, tends to prevent capacitive lumping along the line.

The load capacitance also modifies the propagation delay in the following manner;

$$t_{pd}' = t_{pd} \cdot \mathrm{sqrt}(1 + C_d/C_0\ell)$$

where t_{pd} is the unloaded propagation delay. This increase in propagation delay can also be minimized by a low intrinsic line capacitance and short interconnects. The low dielectric constant (- 3.2) of polyimide produces a lower intrinsic capacitance relative to other types of materials. We have found that a factor of approximately two in speed can be achieved by using polyimide as the dielectric.

Impedance discontinuities can also introduce noise as a result of circuit mismatch along the bus. The low impedance outputs of some driving circuits cause a discontinuity on the line which limits rising edge performance of high-speed busses and introduces reflections. If the electrical length of the line is short compared to the risetime, the reflection from this discontinuity is buried in the risetime and causes no problem; however, if the electrical length is longer, an impedance discontinuity will produce a noise pulse which can introduce extraneous switching of the gates. One should keep the stubs along the line short and evenly distributed to minimize reflective noise. This can more easily be accomplished on multichip packages where distributed points are normally a short distance from the bus line and gates tend to be more evenly distributed.

Present distributed bus architecture is limited due to the RC time constant of the line and the inherent reflections from stubs along the unterminated line. A properly terminated line can minimize reflections; however, terminations are usually undesirable because of the added power generated and the increase in real estate. Our copper/polyimide technology contains relatively low loss lines which provide Gigahertz performance within the package structure. However, for very long lines that sometimes occur between packages, some method of termination must be employed.

Bus lines also consume power due to the charging and discharging of the intrinsic line capacitance. This power is dissipated in the driver circuit. The relationship for CMOS-type driving circuits is;

$$P = 0.5(C_o + C_d)V^2f$$

where f is the frequency of operation and V is the voltage swing. Minimizing C_o by employing shorter interconnects can reduce this power and decrease the size of the driving chip. Honeywell has found that thin film multilayer multichip package technology can reduce the power by a factor of three over technologies presently being used.

INDUCED POWER AND GROUND NOISE: When a large number of the I/O terminals on a VLSI chip switch simultaneously, induced power and ground noise results due to the inductance in the current path. The simultaneous switching noise is given by;

$$V_n = N\ L\ di/dt$$

where N is the number of drivers, L is the inductance path from IC pad to power source, di/dt is the switching current rate.

The problem arises when many I/O's attempt to switch at one time in a particular IC chip. This requires that a large amount of current be supplied instantly to the chip from the power supply via the interconnect inductance path. However, the path inductance will momentarily impede the current flow and cause a current spike that passes through the drivers as signals which can tend to cause erroneous gate switching of other devices on the same signal path. To minimize this inductance, a decoupling capacitor is usually mounted close to the chip power pin. This decoupling capacitor provides the switching current for the chip and decouples the chip from the package trace inductance.

One advantage of the thin film multilayer package is the possibility of placing decoupling capacitors close to the IC pins (some designers are even considering mounting decoupling capacitors on the TAB leadframes). The shorter line lengths within the multilayer package also reduce the inductance path appreciably. In addition, the use of ground planes within the package structure provides a non-inductive, capacitive decoupling function.

DEMONSTRATIONS

The properties of Honeywell's copper/polyimide interconnect system have been well characterized. The electrical properties have been studied and documented through the fabrication of several multichip ECL ring oscillator circuits and transmission line structures (1,2); results have been in good agreement with lumped-element simulations of a lossy transmission line. The thermal properties have been measured and a technique for obtaining a thermal impedance of less than 1° C/W through the polyimide layers has been developed (8). Reliability studies have shown that properly processed polyimide layers can be hermetically sealed in packages to meet the requirements of MIL-STD 883 (9,10). Thus, the copper/polyimide system is now a well established interconnect and packaging technology for present and future electronics systems (4).

Several electronics systems have been packaged at Honeywell. Figure 4 shows a 2.25 inch-square 18 chip substrate mounted in a hermetically-sealable metal flatpack designed for an image processing application (11). Figure 5 shows a control circuit which has a mixture of bare die and surface mount resistors and capacitors for a total of 118 components. This is fabricated on a 180 pinned 3 x 3 inch ceramic substrate with a seal ring. Other circuits have been described in the literature (2-4). A similar gold/polyimide interconnect technology is presently being used in the NEC SX series supercomputers (12).

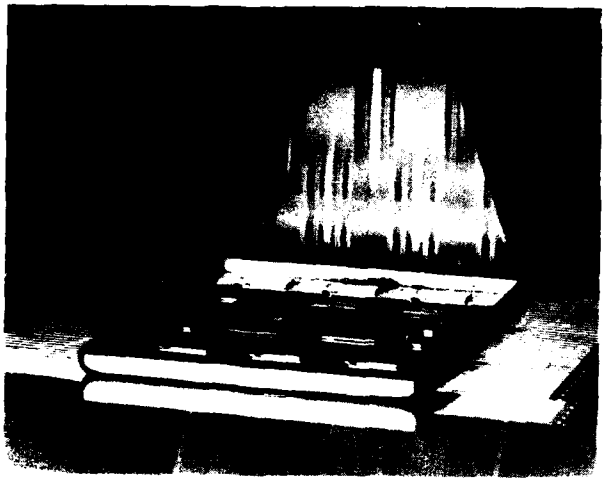


FIGURE 4 An eighteen chip multichip package housed in a metal flatpack fabricated using Honeywell's TFML technology.

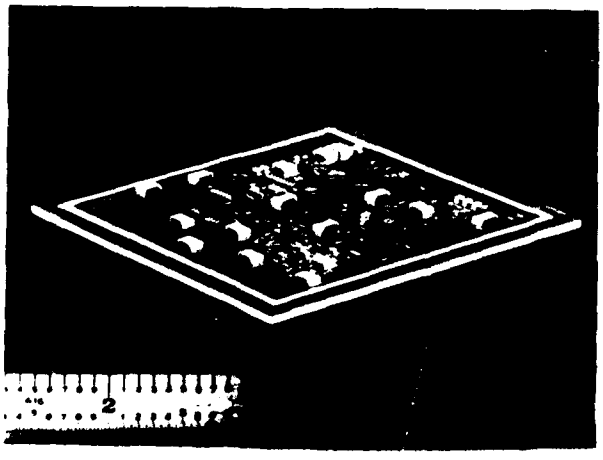


FIGURE 5 Thinfilm multichip interconnects on a pinned, cofired ceramic substrate which interconnects 118 components.

CONCLUSIONS

Multichip packaging approaches are required to interconnect high I/O and high speed devices. Honeywell and other companies have developed thin film multilayer technologies based on copper (or Al, Au, Ni)/polyimide material to fabricate multichip modules. The low dielectric constant of polyimide reduces propagation delay and crosstalk. Use of highly conducting copper minimizes signal loss and risetime slowdown. Thin film IC processing techniques result in fine line geometrics required for high density MCPs. The most fundamental limitation of MCP yield may be posed by the substrate area since large substrates are required to capture functionality in an MCP. In high speed applications, the use of very narrow conductor lines must be limited to short distances to avoid severe signal degradation. The Honeywell approach has large enough lines to be usable at GHz clock speeds for line lengths up to 10 cm. Multichip packaging of VLSICs requires careful evaluation of induced power and ground noise, and the effects of fanout on the busline. The maturity of Honeywell's TFML technology is evidenced by the successful fabrication of several multichip packages on a variety of substrates.

REFERENCES

1. R. J. Jensen, J. P. Cummings, H. Vora, IEEE Trans. Components, Hybrids and Manuf. Technol. CHMT-7, 383-393 (1984).
2. F. J. Belcourt, T. A. Lane, R. J. Jensen, Proc. 37th Electronic Components Conference, 1987, 614-622.
3. R. J. Jensen, "Recent Advances in Thin Film Multilayer Interconnect Technology for IC Packaging," Proc., Electrochemical Society, Oct. 18-23, 1987.
4. R. J. Jensen in Polymers for High Technology: Electronics and Photonics, M. J. Bowden, R. S. Turner, eds., ACS Symposium Series 346, American Chemical Society: Washington D.C., 1987, pp. 466-483.
5. C. J. Bartlett et. al., "Multichip Packaging for VLSI-Based Systems," IEEE ECC, Boston, May 1987, pp. 518-525.

6. C. A. Neugebauer and R. O. Carlson, "Comparison of Wafer Scale Integration with VLSI Packaging Approaches," IEEE Trans. CHMT, Vol. CHMT-10, No. 2, June 1987, pp. 184-189.
7. Raychem Corp., "HDI, High Density Interconnect Design Manual," Document HDI-S 7.4, November 24, 1987.
8. C. J. Speerschneider, F. J. Belcourt, R. J. Jensen, J. M. Smeby, Proc. VHSIC Packaging Conference, 1987, 131-143.
9. R. J. Jensen, R. B. Douglas, J. M. Smeby, T. J. Moravec, Proc. VHSIC Packaging Conference, 1987, 193-205.
10. R. B. Douglas, J. M. Smeby, Proc. 37th Electronic Components Conference, 1987, 197-201.
11. D. J. Kompelein, T. J. Moravec, M. DeFlumere, Proc. Int. Symp. Microelectronics, 1986, 749-757.
12. T. Watari, H. Murano, IEEE Trans. Components, Hybrids, Manuf. Technol. CHMT-7, 462-467 (1985).

ELECTRICAL LAMINATES: WHERE THE INDUSTRY IS GOING AND WHY

Steven J. Kubisen, Parley C. Long
GE Electromaterials Department
1350 So. Second St.
Coshocton, Ohio 43812, USA

ABSTRACT

Electrical laminates can be classified into three general classes. Each class is designed to meet the performance needs of a particular application market. This paper will briefly describe each class and their current performance properties. The improved performance needs will be discussed as well as the new resins and reinforcements being developed to address these needs. Particular emphasis is given to the high performance multilayer products.

SINCE THE ADVENT of the earliest printed circuit boards (PCB's) in the 1950's, new developments in the areas of both active and passive components, as well as overall electronic assembly performance requirements, have determined the direction in which PCB designs and performance capabilities had to evolve. This, in turn, has dictated the requirements which must be met by the laminate material from which these boards are fabricated.

The evolution of the physical size of active components is a good example. Starting with large, bulky vacuum tubes, they evolved through transistors to SSI's, MSI's, LSI's, VLSI's, ULSI's, and GSI's. It has been pointed out that one of the reasons passive components have not followed this trend is the buffering effect of dual in-line packages (DIP's).[1] As long as DIP IC's are in common use, there will not be a great motivation to reduce passive device dimensions to the same extent surface mount device (SMD) technology has today reduced the dimensions of active components. Nevertheless, surface mount devices are inevitably changing the picture. SMD's with several hundred pins and with I/O's approaching 1000 make it quite clear that the largest components, passive devices, are not keeping pace. Eventually, this situation will change, and another significant shrinking of board sizes will follow.

This shrinking of component size and its accompanying need for increased electronic density and speed has created, and will continue to create, major challenges for PC board, and, inter alia, laminate manufacturers. A quick glance at the history of PCB laminate technology will illustrate this point. (Figure 1). Following the emergence of the first PCB's in the 1950's, there followed glass reinforced epoxy technology and NEMA grade G-IO, the forerunner of today's ubiquitous FR-4 laminates. Soon after came the fire retardant FR-2, FR-3, and FR-4's, high temperature epoxy resin systems, polyimides, composite constructions, BT's (bis maleimide-triazines), aramid reinforced systems, quartz fabrics, tetrafunctional and multifunctional epoxies, expanded PTFE's, and most recently, cyanate esters. And, of course, other new high performance materials are continually being developed and brought to the market place. While these developments have been taking place, so, too, has the increased use of multilayer board technology, chip-on-board, tab automated bonding, and assembly improvements such as wave and vapor phase soldering, and automatic insertion of components.

There are several PCB process-related and performance-related requirements that have been dictated by the changes in electronic components and board assembly methods that have occurred during the past thirty years. These requirements include improved copper surface quality, lower dielectric constant substrate materials, improved dimensional stability, predictable material through-process performance, coefficients of thermal expansion (CTE's) that match the CTE's of SMD's, improved machinability, improved flatness, tighter thickness control, improved high temperature tolerance, and, of course, lower cost. These laminate characteristics and their relationship to changes in electronic components and board designs and assembly are outlined below:

IMPROVED COPPER SURFACE QUALITY - Smaller components with increased I/O counts require smaller line and space dimensions. As the width of etched copper conductors and the space between these conductors decreases to meet this demand, defects in the surface copper sheet cause decreased manufacturing yields, thereby raising costs.

LOWER DIELECTRIC CONSTANT - The need for faster electronic signal propagation speeds demands a substrate with a lower dielectric constant. Some designers euphemistically call the ideal substrate "stiff air," referring to a dielectric constant of one.

IMPROVED DS - As line and space dimensions decrease and hole-to-pad ratios increase, any dimensional shift (i.e., expansion or contraction) in the substrate material becomes intolerable from a processability viewpoint. The substrate material must be as dimensionally stable as possible, with the ideal being no movement at all.

IMPROVED PROCESS PREDICTABILITY - If laminate movement cannot be eliminated, the next best thing is for it to move in a predictable way, so that manufacturing processes can be adjusted to repeatably compensate for the movement, thus increasing yields.

LAMINATE CTE = SMD CTE - One of the advantages of using surface mount devices is the resulting improvement in reliability. However, this improvement diminishes as the number of solder joint failures increases. One way to improve solder joint reliability is to match the coefficient of thermal expansion of the laminate material as closely as possible to that of the SMD's. Reinforcement materials such as aramid fiber have proven very effective in providing a way to match these CTE's.

MACHINABILITY - As board feature sizes decrease, drilled hole diameters decrease and overall dimension tolerances become tighter. Therefore, new materials must be better able to be drilled, punched, and routed.

IMPROVED FLATNESS - Several aspects of PCB performance demand a laminate material that can remain extremely flat throughout all board fabrication, assembly, and test steps. Critical are the ability to screen print very small surface features, accurate photoimaging capability, automatic component placement, wave soldering at high temperatures, and "bed-of-nails" electrical testing.

TIGHTER THICKNESS CONTROL - This is an outgrowth primarily of controlled impedance board designs. Excessive variations in layer-to-layer thicknesses can cause intolerable variations in impedance between adjacent conductors within the board.

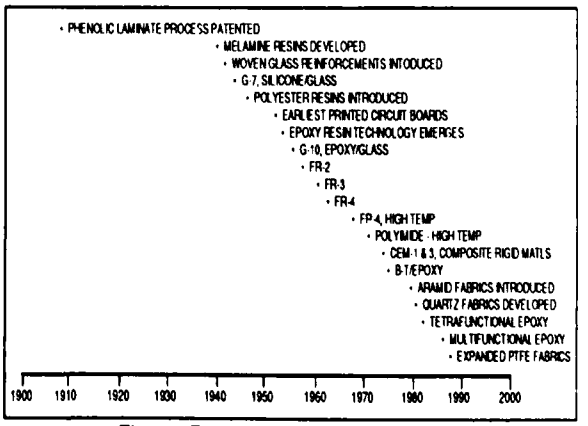


Fig. 1 - Electronic Substrate Timeline

HIGHER TEMPERATURE TOLERANCE - PCB's are today being exposed to more, and more severe, temperature extremes than ever before, and indications are that the situation will only become worse in the future. Temperature related problems result from two environmental conditions: (1) very high absolute temperatures, and (2) repeated cycling between different temperatures. Because the Z-axis CTE of most dielectric materials is greater than the CTE of copper, thermal cycling fatigues through-hole plated copper, leading to barrel cracking and, therefore, board failure. Very high absolute temperature cause higher overall Z-axis expansion, which, again, causes plated-through-hole barrel cracking. (Repeated repairing of the board combines high temperatures with temperature cycling.) One way to minimize these problems is to utilize a laminate material with as high a glass transition temperature (Tg) as possible.

LOWER COST - Obviously, the realities of business dictate that the cost-to-performance ratio of a PCB be as small as possible, to maximize profitability. Globalization of the electronics industry has severely exacerbated this problem.

The effect these trends have had on laminate development and usage is clearly reflected in Figure 2. Despite the desires of the market place, with increased performance has, inevitably, come higher cost. This trend, as it relates to laminate material, is shown in Figure 3.

General Electric conducted a blind mail survey of the US electronics

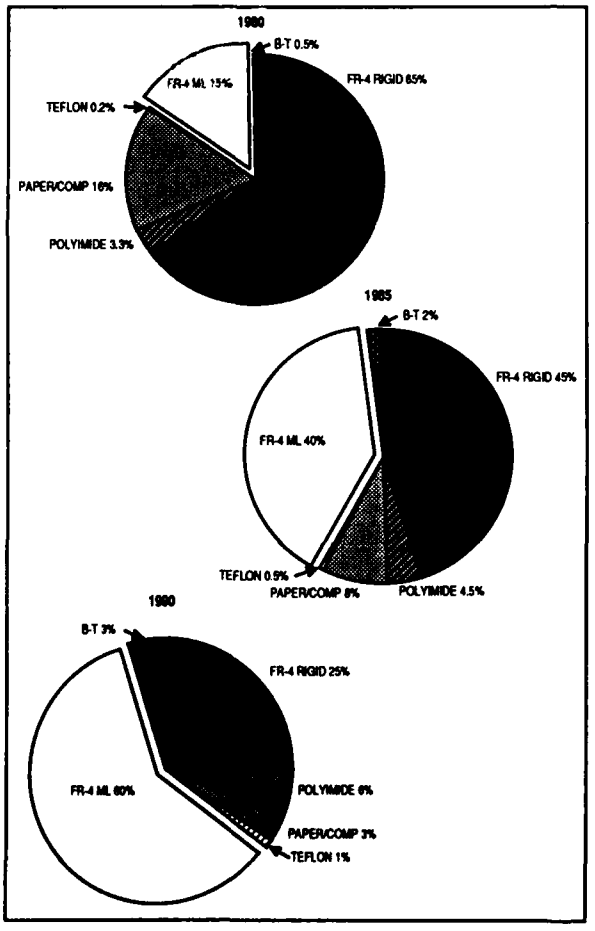


Fig. 2 - Laminate Usage

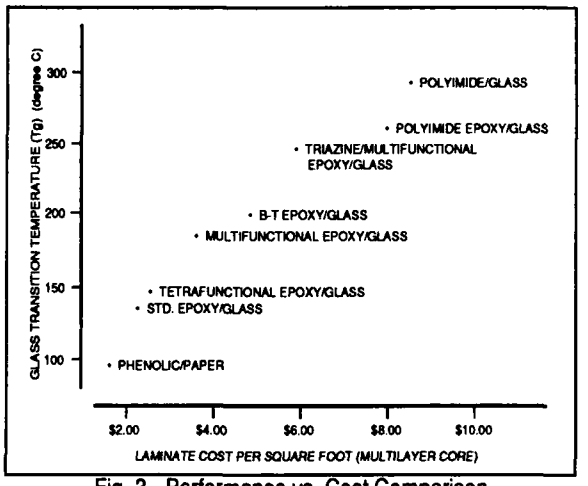


Fig. 3 - Performance vs. Cost Comparison

industry, targeting those individuals who buy and/or specify printed circuit laminate materials. Three questions are of interest to our discussion. First, what did those surveyed believe to be the limitations of today's materials. The most frequent limitations were:

Tg TOO LOW	23.5%
POOR DS	11.4%
DIELECTRIC CONSTANT TOO HIGH	7.3%
Z-AXIS EXPANSION IS TOO GREAT	5.7%
EXCESSIVE WARP & TWIST	4.9%
COST IS TOO HIGH	4.3%
X-Y CTE DOESN,T MATCH SMD'S	4.1%
POOR QUALITY	3.0%

Clearly, Tg and DS stand out as major areas of concern.

The second question was, what did those surveyed believe today's "ideal" substrate would be. The answers:

FR-4 IS OKAY	15.0%
HIGHER Tg	10.4%
LOWER DIELECTRIC CONSTANT	8.7%
"POLYIMIDE"	6.8%
LOWER COST	6.0%
X-Y CTE MATCHES SMD'S	5.5%
"CERAMIC"	4.4%
LOWER Z-AXIS EXPANSION	3.3%

Again, a higner Tg is viewed as very important to improving board performance.

The third question, what would it take to meet the needs of the industry five years from now, elicited the following responses:

HIGHER Tg	21.7%
IMPROVED DS	19.1%
LOWER DIELECTRIC CONSTANT	19.1%
LOWER COST	13.2%
LOWER Z-AXIS EXPANSION	8.6%
X-Y CTE MATCHES SMD'S	7.2%

A higher Tg material with improved DS performance and a lower dielectric constant seems to be needed both now and in the future.

The current market for electrical laminates is composed basically of three primary classes of laminates. These include paper reinforced laminates, composite laminates (containing woven glass and a nonwoven reinforcement), and woven glass reinforced laminates. Other woven reinforcements are also used, but to a much smaller extent than woven glass.

Paper laminates are based on phenolic resins. Common UL/ANSI grades of paper laminates include XPC, XXXPC, and FR-2. Typical properties for FR-2 are listed in Table I. Paper based laminates are commonly used in consumer electronics applications. The low Tg and low flexual strength and flexual modulus limit their use in applications involving exposure to elevated temperatures and applications involving high component loadings. The high z-axis expansion and high moisture absorption also prevent their use in applications involving plated thru holes (PTH).

Table I - Electrical Laminate Properties

Property	FR-2	CEM-1	FR-4
Thermal			
Tg (degree C)	80	90	125-135
Z-axis Expansion (23 to 260 degree C, %)	N/A	6.3	4-5
Electrical			
Dielectric Constant (1MHZ)	4.5	3.9	4.5-5.0
Dissapation Factor (1MHZ)	0.024	0.028	0.020-0.030
Dielectric Strength (V/mil)	500	600-800	750-1000
Physical			
Flexual Strength (PSI)			
— Lengthwise	12,000	50,000	75,000
— Crosswise	10,000	36,000	58,000
Flexual Modulus (PSI)			
— Lengthwise	$2.1x10^4$	$2.4x10^6$	$2.8x10^6$
— Crosswise	$2.0x10^4$	$2.0x10^6$	$2.5x10^6$
Peel Strength (lb/in)	6-2	8-10	9-11
Water Absorption (%)	.6	.10-.15	.1
Solder Float (sec. @ 550 deg. F)	30*	25-40	>60
Flammability	V-0	V-0	V-0

* Value for 500 degree F

Composite laminates are based on epoxy resins. Common UL/ANSI grades of composite laminates include CEM-I and CEM-3. CEM-I is constructed from woven glass face sheets and a cellulose paper core. CEM-3 is constructed of woven glass face sheets and a nonwoven glass core. Typical properties for CEM-I are reported in Table I. Composite laminates are commonly used in consumer electronic applications. They offer improvements in flexual strength and flexual modulus over paper laminates which provide the ability to achieve higher component loading on the circuit board. However the low Tg and high z-axis expansion also preclude its use in PTH applications.

Woven glass reinforced laminates are by far the most common electrical laminates. The FR-4 grade is based on an epoxy resin system. The common epoxy resin used in FR-4 is a difunctional epoxy based on the reaction of the diglycidyl ether of bisphenol A and tetrabromobisphenol A. The resin is typically crosslinked with a dicyandiamide hardner and is catalyzed by an amine catalyst. Typical properties for FR-4 are reported in Table I. The higher Tg and lower Z-axis expansion allow FR-4 to be used in PTH and surface mount device (SMD) applications. However, as we have already discussed, even the property profile of FR-4 is becoming insufficient for the increasing requirements being placed on the electrical rigid and multilayer laminates.

Higher Tg resin systems do offer improvements in resistance of the laminate to property degradation during exposure to high temperatures in processing or in the final applicaton. A higher Tg resin system will also yield a lower CTE needed for PTH reliability and SMD reliability. The CTE of a material is much less below its Tg than above its Tg. For a typical thermoset resin the CTE below its Tg is 60 - 70 ppm/ degree C and 300 ppm/degree C above its Tg. Thus by increasing the Tg of the resin, one can decrease the effective CTE of the laminate over the typical temperature range seen in processing (23 to 260 degree C).

One approach to increasing the Tg of the typical FR-4 epoxy resin system is to increase the crosslink density of the resin system. Multifunctional expoxies (trifunctional epoxi[illegible]functional epoxies) are used in the epoxy resin system to increa[illegible]x[illegible]v and [illegible] increase Tg. Typical properties for multifunctio[illegible] epoxy la[illegible] are [illegible]ed in Table II. The Tg's of these resin syste[illegible] increased 10 to 30[illegible]rees C above that of difunctional epoxies. The [illegible] expansion is also reduced from 4-5% to 3-4%. Other properties such as dielectric constant, dimensional stability (DS), and flatness are not affected the multifunctional epoxies.

Table II - Advanced Laminate Properties

Property	Multi-Function	BT/Epoxy	Polyimide
Thermal			
Tg (degree C)	130-160	175-190	250-280
X-axis Expansion (23 to 260 degree C, %)	3-4	2-3	1-2
Electrical			
Dielectric Constant (1MHZ)	4.5-5.0	4.1-4.8	4.0-4.7
Dissapation Factor (1MHZ)	.020-.030	.13-.020	.010-.020
Dielectric Strength (V/mil)	800-1100	1000-1500	1000-1500
Physical			
Peel Strength (lb/in)	8-11	7-9	6-8
Water Absorption (%)	.1	.3	.6
Flammability (UL 94)	V-0	V-0	V-1

Another approach to increasing the Tg of a difunctional epoxy system is incorporation of another high Tg resin such as a triazine resin. Triazine resins are based on the trimerization of cyanate monomers. These resins offer a high Tg and good electrical properties; however, they are reported to suffer from poor water resistance at high temperatures.[2] Incorporation of a bismaleimide monomer with the triazine monomer improves the water resistance of the system. These BT resins (bismaleimide-triazine) are commonly blended with epoxy resins in systems used commercially. Typical properties for BT/epoxy resin systems are reported in Table II. The Tg of these systems is raised to the 175 - 190 degrees C range with a corresponding reduction in the Z-axis expansion to 2-3 %. There is some reduction in

the dielectric constant of the system, but the copper peel strength is also reduced. The DS of BT/epoxy systems is reported to to improved over conventional FR-4 resins since the resin shrinks less on curing.[2] No improvements in flatness are reported for BT/epoxy systems. Because BT/epoxies do involve the addition of a nitrogen containing resin to the system, the cost of the resin system is relatively high.

One of the highest Tg polymers know is polyimide. As reported in Table II, polyimides offer Tg's as high as 280 degrees C. These polyimide systems offer Z-axis expansion values of 1 - 2 %. and offer some improvement in dielectric constant (i.e. lower). The polyimide systems do not yield copper peel strengths as high as those of FR-4 and are suseptible to moisture absorption. In addition, due to the extremely high Tg and brittle nature of the polymer system, polyimides exhibit no drill smear but do need to be drilled at lower chip loads to prevent chipping of the hole walls. The original polyimide systems were based on methylene dianiline (MDA). MDA has been reported as an animal carcinogen and a possible human carcinogen. These initial systems also were only V-I in the UL94 flammability test, whereas most applications require a more stringent V-0 rating.

More recently MDA free polyimide systems have been introduced.[3] These systems are blends of the MDA free polyimide and a multifunctional epoxy. These systems offer slightly higher Tg's (290-310 degrees C) and also yield V-O in the UL94 flammability test. These newer systems are also suseptible to moisture and do require a plasma.etch system for hole cleaning.

LOW DIELECTRIC CONSTANT MATERIALS - The lowest dielectric constant resin currently in use for electrical laminates is polytetrafluoroethylene (PTFE). PTFE has a dielectric constant of 2.1 and a Tg of 75 degrees C.[4] This low Tg limits the use of PTFE to applications which must have the extremely low dielectric constant such as high frequency microwave.

The other approach to reducing the dielectric constant of the laminate is to choose reinforcement materials which offer lower dielectric constants. Table III reports the dielectric constants of the commonly used E-glass and other fibers which will offer improvents in dielectric constant.

Table III - Dielectric Constants of Reinforcements

Material	Dielectric Constant
E-Glass	6.5
S-Glass	5.2
D-Glass	3.8
Quartz	3.8
Aramid	4.1

LOW CTE MATERIALS - For critical surface mount devices the CTE of the laminate in the X-Y plane is very important. The ceramic chip carrier typically has a CTE of 5 - 7 ppm/degree C vs a typical glass epoxy laminate value of 16 ppm/degree C. The approach the laminate industry is taking for these critical applications is to utilize low CTE reinforcements or metal restraining layers.[5] Using a woven aramid fiber in combination with a tetrafunctional epoxy resin will yield X,Y CTE values of 5 - 7 ppm/degree C. Metal restraining layers such as copper-Invar-copper or Alloy 42 are traditionally used as a base material for substrate materials or as an internal power or ground layer in multilayer assemblies. The X,Y CTE's of these various reinforcement materials are reported in Table IV.[6,7]

Table IV - X,Y Coefficient of Thermal Expansion of Reinforcements

Material	CTE (PPM/Degree C)
E-Glass	5.4
S-Glass	2.6
Quartz	.5
Aramid	(2.0)
Cu-Invar-Cu	5.0
Alloy 42	5.0
Molybdenum	4.9
Copper	17.6
Alumina	6.4

NEW LAMINATION PROCESSES - In addition to the new high Tg resins and improved reinforcements, new process methods are also being developed.[8] Conventional laminates are manufactured in a multi-opening "daylight" press. Typically 8 to 10 sheets of laminate are pressed in each opening. As the heat is applied to the laminates, the inner laminates see a different cure profile compared to the outer laminates due to the thermal lag in the system.

The new lamination process which is being developed involves a continuous lamination. In this process each laminate is exposed to exactly the same cure profile. This uniform cure profile is reported to yield laminates with improved dimensional stability and improved thickness uniformity across the laminate. Due to fewer handling steps, improvement in the copper surface is also anticipated.

From the above discussion it should be quite clear which direction the electronics industry is going, and what characteristics the ideal laminate material of tomorrow should possess in order to meet future board demands. Briefly, the laminate of the future should have a high Tg, say, over 200 degree C; it should have excellent dimensional stability, with the ideal being no movement at all; it should have, for some (but not all) applications, a low dielectric constant; it should have a CTE approaching that of surface mount devices; and, last, it should be inexpensive.

Will the laminate industry keep pace with the demands placed upon it by the electronics industry? Undoubtedly it will, but probably not at the pace board designers and fabricators would like. One thing, however, does seem certain. The laminates that will meet the challenges of tomorrow have yet to leave the laboratories of today.

ACKNOWLEDGEMENT

The authors would like to thank A.R. Senese and J.E. Tracy of GE Electromaterials Department for their assistance and helpful suggestions in preparing this paper.

REFERENCES

1. Hartley, N., "Yield Problems in SMT and Development of Process Equipment", IPC Technical Review, March-April 1988, pp. 14-23

2. Gaku, M., "Resin Laminate Resists High Heat", Circuits Manufacturing April 1980, pp. 53-57

3. Davis, M.J., Senese, T.R., "A New Polyimide for Multilayer Laminate Applications", IPC Fall Meeting, Chicago, October 1987.

4. Wellig, L.R., "Coming Soon: The Ideal Laminate", Circuits Manufacturing, January 1988, pp. 25-28

5. Cairns, G.R., "Controlled CTE Materials", IPC Fall Meeting, Chicago, October 1987

6. Dance, F.J., "Interconnecting Chip Carriers: A Review", Circuits Manufacturing, May 1983

7. Capiollo, C., "How to Design Reliability into Surface Mount Assemblies", Electronic Packaging & Production, July 1985

8. Eliasson, R., "How to Make Better FR-4", Electronic Packaging & Production, October 1987, pp. 76-77

THERMAL EXPANSION MISMATCH AND FATIGUE IN ELECTRONIC PACKAGES

Lewis S. Goldmann
IBM East Fishkill Facility
Hopewell Junction, New York, USA

ABSTRACT

Electronic packages, comprised of dissimilar materials and subjected to repeated thermal cycles, are potential candidates for thermal mismatch fatigue. A survey will be presented of the analytical and experimental tools available to evaluate a package's susceptibility to this mode of failure, with an emphasis on the recent literature.

ALL ELECTRONIC PACKAGES are comprised of dissimilar materials and virtually all experience periodic temperature excursions. This simple combination of facts make almost all packages potential candidates for failure from thermal fatigue. If the structure is improperly designed, fabricated or utilized, the repeated imposition of a thermal expansion mismatch between elements of the package can result in premature mechanical failure of a component, resulting in possible loss of functionality.

With the increased emphasis on miniaturization in recent years, the threat of thermal fatigue has been magnified in two ways. First, there is greater power dissipation per unit volume which has not been fully matched by thermal management, thus giving rise to greater temperature excursions. Secondly, reduction in size of the interconnections which join the heat-generating circuits or devices to the computer or microprocessor can result in an increased stress level.

Reliability and design engineers in the electronic industry are becoming increasingly aware of the thermal mismatch problem, and have at their disposal a wide array of analytical and experimental tools for product optimization and characterization. The literature has expanded enormously, with hundreds of papers now appearing annually. These vary from basic material, stress and fatigue studies, to the design, analysis and testing of specific package applications.

The literature, and the entire field, may be roughly divided into four categories, and as a brief introduction to the field, this survey will follow the same subdivision:

System displacement and forces.
Stress analysis of critical elements.
Accelerated testing.
Fatigue prediction.

The complete evaluation of an electronic package (i.e. starting with an existing package design or prototype, prediction of functional fatigue lifetime or failure rate under the actual operating environment) involves all four subjects, but their complexity and especially the seams between them has thus far prevented this ideal from being achieved. Nevertheless, the more of these tools that are utilized, individually or in combination, the greater the likelihood that the package will successfully meet its lifetime requirements with no degradation or loss of functionality. For greater detail on the following subjects and a more extensive bibliography, the reader is referred to Ref. [1].

SYSTEM DISPLACEMENTS AND FORCES

Usually the most susceptible elements to thermal fatigue are the interconnections, or joints, between the chip

and its carrier, or among the additional levels of packaging to the outside world. These experience repeated stress due to the thermal expansion pattern of the entire package. Thus the starting point in understanding a potential thermal fatigue situation is characterization of the whole package or system. For instance, if a wire bond is suspect, the expansions and interactions of chip, substrate and encapsulant, as well as of the wire itself, all come into play. Determination of the system, or global, displacements and forces can be considered a macro-evaluation, providing the input conditions for the stress study of the critical interconnection, which can be thought of as a micro-evaluation.

The underlying principle of the system evaluation is simply that the entire mismatch in thermal expansion between any two elements must be accommodated by mechanical deformation of the elements themselves and of any intervening connections. Allocation of these deformations depend upon the relative mechanical stiffness of the various components.

System evaluation may be approached by a wide variety of techniques, both experimental and analytical, often in combination. Beacuse the driving stress parameter is invariably temperature, the evaluation usually involves measurements or modeling which are thermal as well as mechanical.

Recently, much attention has been given to the Surface Mount Technology (SMT) structure of a leadless ceramic chip carrier (LCCC) bonded to a printed wiring board (PWB) by a multiplicity of solder joints*. Conceptually this package is nearly identical to a flip-chip on ceramic substrate. The large expansion mismatch between the LCCC and PWB severely deforms the interconnections. If the joints are assumed to offer no resistance to LCCC and PWB expansion, they are subjected to a simple shearing motion [3,4]. Often, however, the solder joints are sufficiently numerous and/or rigid to constrain free expansion, resulting in mechanical LCCC deformation, both in-line and flexural [1,5,6], as shown schematically in Fig. 1. Moreover, the chip or LCCC may have its own natural thermal curvature because of dissimilar components or films. For simple, peripheral arrays of joints, analytical expressions are available to describe these complexities [1], but more generally, complete understanding of system deformation requires experimental measurement, such as by strain gage [5]; or finite element modeling (FEM).

System deformations of leaded structures, such as dual in-line packages (DIP), where leads are soldered through PWB's, or leaded chip carriers, may be evaluated in the same manner as leadless carriers. These rigid leads will almost always cause carrier or module deformation by constraining free expansion.

In cases where the rigidity of interconnections is known to induce mechanical deformations of the carrier, joint stiffnesses must be estimated, analytically or experimentally, in each of the pertinent deformation modes. Considering the gull wing lead shown in Fig. 2, the stiffness matrix must be determined which relates horizontal and vertical displacement and rotation to horizontal and vertical force and bending moment.

The role of interior* polymer encapsulants in thermal mismatch reliability can be very important, both in terms of mechanical effects (of interest here) and mechanical-chemical synergisms. Because of their complex material properties and generally irregular geometry, an exact analytical approach is rarely viable. System evaluation is generally made by approximate closed-form analysis [7-9], photoelasticity [10], FEM [8] or in situ strain gages [7,11,12]. Any evaluation which requires elastic and thermal expansion properties of a polymer should, if rigorous, take into account its sensitivity to temperature, strain rate and history. The usual short-cut is to assume a single, nominal elastic modulus and thermal expansivity, which is generally adequate unless the temperature range of interest is near the glass transition, where the modulus may change by several orders of magnitude.

STRESS ANALYSIS OF CRITICAL ELEMENTS

The cyclic or repeated stress on an interconnection or other critical structural element is a function of the system displacements and forces it experiences, and of its own geometric and materials characteristics. Consider the unencapsulated gull-wing lead of Fig. 2 and assume it is soldered to a PWB as shown in Fig. 3. The lead applies a system of forces and moments to the solder joint which sets up an internal stress field.

---

*An excellent survey of SMT solder joint fatigue may be found in Ref. [2]

*We refer here not to structural polymers, such as PWB's or molded DIP casings but to the softer protective encapsulants used inside a package. The former must of course be treated as any other structural element in the System Evaluation.

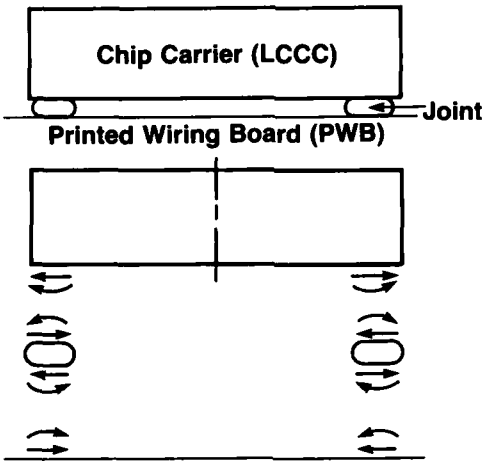


Fig. 1. Schematic of forces and moments generated in an LCCC (low expansivity) and PWB (high expansivity) during heating.

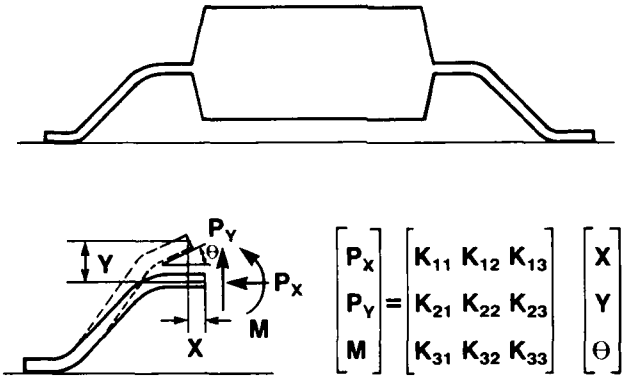


Fig. 2. Stiffness matrix of a gull-wing lead.

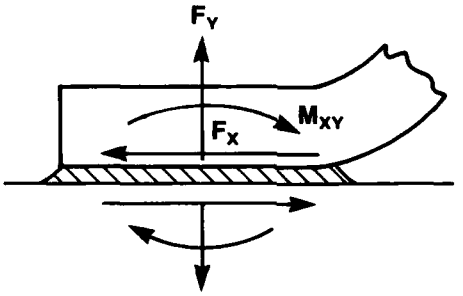


Fig. 3. Forces transmitted to solder joint from gull-wing lead.

The more severe the stresses the more likely a fatigue problem will occur. If a crack does develop, it very often is located at or just adjacent to an interface, where metallurgical imperfections and high concentration gradients give rise to stress concentrations.

While thermal stresses in interconnections are most frequently modeled by FEM [13-17], approximate closed form analyses (of solder joints [1,4,5], die bonds [18], castellations [19], plated through holes [20], etc.) have been published, as have been exact solutions for idealized structures (spherical-segment solder joint modified to a right circular cylinder [21], two-dimensional die-bond [22,23], etc.).

All of the above references assume that that the materials are linearly elastic. Unfortunately, solders, which are universally used with surface mounted as well as pinned components and hence the focus of most recent work in the field, exhibit significant plastic flow, even at moderate stress levels and temperatures. While many FEM codes can accept non-linear properties, the extent of the experimentation required to determine these properties have until recently precluded non-linear FEM of interconnections. A recent paper [24] describes an elastoplastic three-dimensional FEM of solder castellations, but in the absence of measured properties assumes classical forms for the plastic flow relationships. An approximate non-linear analytical solution of an axisymmetric solder joint has been published [1], which employs simplifying assumptions of both deformation pattern and constituitive equations.

The stress situation for solders is further complicated by the slow heating and cooling rates and long dwell times typical of electronic component applications. Operating at an absolute temperature over half their melting point, solders experience significant creep and stress relaxation, which, ultimately, must be taken into account in rigorous modeling. Considerable data has been published on solder creep and relaxation, usually on bulk specimens under uniaxial stress [25,26,27]. More recently, in situ solder joint response to an actual thermal cycle environment has been quantitatively inferred from strain gage measurements on an LCCC [5], and attempts made to incorporate the data into an approximate stress model [28].

Thus, with the analytical tools and problem formulation in hand, and with the foundation of an appropriate data base, pursuit of the ultimate, time-dependent solder interconnection stress analysis will continue to prove a fertile field for investigators, with further advances likely to appear in the near future.

ACCELERATED TESTING

Faced with the difficulties in estimating thermal fatigue lifetime purely by analytical means, reliability and design engineers have turned to accelerated testing to augment or replace modeling. In accelerated testing, some critical parameter, such as strain, frequency or temperature is exaggerated in order to reduce the number of cycles, or test time, to product failure. Then, by mathematical means, the results are extrapolated to the actual operating environment of the component. As thermomechanical understanding of electronic packages, and of interconnections in particular, have improved, likewise have the sophistication and specificity of testing. As each package has its own distinctive structure and material set, it is clear that accelerated testing should be done on the actual product, or on a structural clone modified to provide electrical measurement capability.

Single-cycle test measures, such as strength or ductility, have been shown to have value as fatigue predictors [29,30] for bulk materials. It is very dangerous, however, to apply this principal to the complex structures and material sets that comprise today's electronic packages [14]. For instance, a pull test may be routinely used to monitor the quality of wire bonds, but should not be used to predict fatigue lifetime, as it may rupture an entirely different interface.

Mechanical fatigue cycling of interconnections [31,32,33] in a mode which replicates system thermal mismatch is commonly, though often incorrectly, used. It is most applicable for metals which will be used at temperatures below half their absolute melting point, such as copper conductors or gold wires; and when the amplitude of the test parameter (force, deflection, rotation, etc.) can be related to the corresponding operating amplitude, taking into account system displacements. Mechanical cycling of solder joints requires knowledge of time- and temperature-dependent effects in all their complexity, without which extrapolation to operating conditions is very tenuous. Nevertheless, the rapidity in collecting cycles and ease in obtaining strain amplification often make mechanical cycling the most attractive test mode, and, recognizing this, recent work [34,35] has emphasized measurement and accommodation of creep and other time-

dependent properties in mechanical cycling of solder joints.

Thermal cycling of hardware in controlled-profile chambers [5,9,16,36, 37] is the most widely used acceleration test for mismatch fatigue of electronic packages. In contrast to mechanical cycling, it exposes the package to a combined thermo-mechanical cycle, and it reflects, to a substantial degree, the mismatch subdivision among the various elements. Its major shortcoming is the inability to simulate thermal gradients experienced by the package in the operating environment. One effect of gradients is alteration of the overall system mismatch, both during power-up (which may be the more severe [38]) and at steady-state. By judicious extrapolation from accelerated to operating conditions, however, this discrepancy can be mathematically accommodated [1]. Another potential manifestation of gradients - warping of a chip or carrier - is not easily compensated for [39]. If an electronic component is known to warp significantly in actual operation, then its package is probably a good candidate for power cycling. Another application which calls for power cycling is a back-bonded power device. It has been shown [40] that voids adjacent to the substrate may lead to thermal runaway and catastrophic bond fracture in operation, but that the phenomenon would not appear in thermal cycling.

In power cycling [39,40,41,42], the actual device, or a structurally identical but specially wired replacement, is repeatedly powered to a controlled level. While demanding more sophisticated equipment and controls than thermal cycling, power cycling is often the closest simulation to the actual machine environment of an electronic package. It neglects only metallurgical and chemical time-dependent factors such as corrosion and solder grain growth. There are occasions, however, when thermal analysis shows that the temperature history of a particular component will depend more on heat transfer from the machine than from internal generation. In these cases thermal cycling is probably the better choice.

Regardless of whether mechanical, thermal, or power cycling is used, two requirements must be fulfilled: there must be acceptable methods of detecting the end of life, and of extrapolating the results to actual operating, or field, cycles. The latter will be discussed in the following section.

Most commonly, an electrical criterion is used to detect accelerated fatigue failure, either four-point joint resistance, or circuit functionality, of discrete joints. Resistance is usually not sufficiently sensitive for large area interconnections, such as die bonds, and is replaced by visual crack observation, thermal resistance [40,42], or ultrasonics [37]. With mechanical cycling, a decrease in mechanical resistance to deformation is diagnostic of a fatigue crack.

FATIGUE PREDICTION

Proper choice of test parameters and accurate calculation of acceleration factors are vital for reliability assessments. Extrapolation to the operational environment requires use of an acceptable fatigue model.

The frequently use Coffin-Manson equation [29], which relates fatigue life (N_f) to the elastic and plastic (ϵ_p) components of strain amplitude, works well for metals which do not exhibit creep, relaxation or other time/temperature dependencies. When the plastic strain predominates, the equation simplifies to

$$N_f \propto \frac{1}{\epsilon_p^2} \qquad [1]$$

If this equation were valid, the acceleration factor between test and actual cycles would simply be the square of the strain amplitude ratio. Moreover, if the joints exert zero constraint, or if the entire system is linearly elastic, then strain amplitude is directly proportional to the temperature amplitude of cycling, and the strain need not be calculated. In the more general case, identification of a meaningful strain amplitude in Eq. 1 should be related back to the stress analysis of critical elements. Often a simplified parameter, such as gross shear strain, is used.

This simple Coffin-Manson equation has been and continues to be used for many studies of solder interconnections [30,42,43,44]. However, in the last two decades, it has become increasingly clear from bulk solder fatigue studies (eg.[45-48]) that Coffin-Manson is inadequate to describe the behavior of materials which are cycled at temperatures above half their absolute melting point. When significant creep is involved, even the "invariant" strain exponent in eq. (1) is on shaky ground [38,44,49]. Many alternative formulations have been proposed, but to date none predominate. Among the approaches are: (a) modifying eq. (1) with empirically derived terms involving cyclic frequency and maximum temperature [38,50,51]; (b)

strain-range partitioning, which weighs both creep and plastic flow contributions in a Coffin-Manson type formulation [52,53]; (c) energy dissipation considerations [47,54]; (d) crack propagation algorithms [55,56,57]; and (e) other quasi-empirical approaches (eg. [46]). These various approaches have utilized a vast body of experimental materials data, primarily bulk specimens in isothermal, uni-axial fatigue or in long-term stress relaxation. Recently, however, as the complex effects of cycle conditions and terminal geometry have become more widely recognized, testing associated with the development of analytical models has begun to more closely reflect operating conditions and to be applied to product-like interconnection geometries [5,34,35, 53].

Accelerated fatigue data, modified by an appropriate acceleration factor, produces a projection of product lifetime in an operational environment. Normally this average or mean lifetime is inadequate for reliability estimates. What is needed instead is a value for failure rate at any prescribed lifetime. For this, one needs not only a mean, but a distribution of lifetimes, implying the need for a sufficiently large sample size in the accelerated test to insure statistical confidence in the distributional parameters. Discussion of lifetime statistics are beyond the scope of this survey.

LIFETIME EXTENSION

Underlying the rapid growth of interest in thermal mismatch fatigue is the need to maintain high product reliability in the face of increased device density, larger chips, decreased interconnection size, and new structural concepts. With greater understanding have come numerous schemes to reduce the risk of thermal fatigue which involve modifying the package design or materials.

Significant advances have been made in minimizing the stress on critical components. For instance, a PWB fabricated of a copper-invar sandwich [58,59], or laminated of a low expansivity organic such as Kevlar [60-62], has a much closer expansivity match with an LCCC than the traditional epoxy-glass. Interconnection array footprints may be modified to reduce the distance from the furthest joint to the stationary or neutral point. Compliant surface-mounted or pinned leads may be designed for greater strain relief, allowing them to absorb a greater percentage of the overall mismatch.

The shape of the interconnection may be designed to minimize local thermal strains. In the case of solder, this usually entails making the joint taller and more slender [13,63-65], or stacking joints [66]. A summary of recent progress in the lifetime extension of solder joints may be found in another paper in these proceedings [67].

Proper choice of the interconnection material can improve fatigue lifetime. Die bonds, for instance, traditionally have been [42,68] either soft (lead-based solders, filled polymers) or hard (gold-based eutectics, filled glass [69,70]). The former offer little resistance to chip expansion and are thus susceptible to thermal fatigue [37,40,42,71]; the latter constrain the die and are thus themselves fatigue resistant, but could cause die fracture [72,73,74]. Bridging the gap, intermediate materials, neither hard nor soft, have been developed [42] to maximize overall product reliability. For SMT joints, alternative solders to the traditional lead-tin - notably lead-indium - have shown longer thermal cycle lifetime [75,76], with positive results demonstrated for plated through holes as well [77].

CONCLUSION

Thermal mismatch fatigue in electronic packages has become an extremely active field in the last decade, stimulated to a great extent by the advent of surface mount technology. Augmented by the inevitable trends to greater device density and smaller size, this interest is likely to continue to grow in importance and activity level in the forseeable future. Much remains to be done in each of the four categories described above; and especially in regard to their merging into a single, coherent reliability assessment.

REFERENCES

1. R.R. Tummala and E.J. Rymaszewski, editors, "Microelectronics Packaging Handbook", to be published by Van Nostrand, 1988: Chapter 5 (Reliability) by D.A. Jeannotte, L.S. Goldmann and R. T. Howard, Section 5.5.

2. Lau, J.H. and D.W. Rice, Solid State Technology, Oct. 1985, 91-104.

3. Goldmann, L.S., IBM Journal of Research and Development 13(3), 251-65 (1969).

4. Ohshima, M., A. Kenmotsu and I. Ishi, Proc. IEPC 1982, 481-8 (1982).

5. Hall, P.M., IEEE CHMT-7(4), 314-27 (1984).

6. Hall, P.M., T.D. Dudderar and J.F. Argyle, IEEE CHMT-6(4), 544-52 (1983).

7. Usell, R.J. Jr. and S.A. Smiley, Proc. Reliability Physics Symp., 65-73 (1981).

8. Baker, E., IEEE PMP-6, 121-8 (1970).

9. Uebbing, J.,Proc. Reliability Physics Symp., 149-56 (1981).

10. Liechti, K.M., Experimental Mechanics, Sept. 1985, 226-31.

11. Dale, J.R. and R.C. Oldfield, Microectronics and Reliability, 16, 255-8 (1972).

12. Schroen, W.H., J.L. Spencer, J.A. Bryan, R.D. Cleveland, T.D. Metzgar and D.R. Edwards, Proc. Reliability Physics Symp., 81-7 (1981).

13. Wilson, E.A. and E.P. Anderson, Proc. 33rd Electronic Components Conf., 320-7 (1983).

14. Sherry, W.M., J.S. Erich, M.K. Bartschat and F.B. Prinz, Proc. 35th Electronic Components Conf., 81-90 (1985).

15. Walker, D.L., L.R. Fox and R.J. Hannemann, IEEE CHMT-6(3) (1983).

16. Smeby, J.M., Proc. 34th Electronic Components Conf., 117-24 (1984).

17. Clatterbaugh, G.V. and H.K. Charles, Jr., Proc. 35th Electronic Components Conf., 60-72, (1985).

18. Taylor, T.C. and F.L. Yuan, IRE Trans. Electronic Devices ED-9, 303-8 (1963).

19. Riemer, D.E. and J.D. Russell, Proc. ISHM 1983, 217-22.

20. Frankland, H.G., N.J. Sawyer and I.G. Sanderson, Insulation/Circuits, Jan. 1971, 21-7.

21. Robert, M. and L.M. Keer, Quarterly Journal of Mechanics and Applied Mathematics 40(3), 365-81 (1987).

22. Chen, W.T. and C.W. Nelson, IBM Journal of Research and Development 23(2), 179-88 (1979).

23. Zeyfang, R., Solid State Electronics 14, 1035-9 (1971).

24. Lau, J.H., D.W. Rice and P.A. Avery, IEEE CHMT-10(3), 346-57 (1987).

25. Ellis, J.R. and E.P. Esztergar, ASME Symposium on Design for Elevated Temperature Environment, 29-43 (1971).

26. Baker, E. and T.J. Kessler, IEEE PHP-9(4), 243-6 (1973).

27. Baker, E., Material Science and Engineering, 38, 241-7 (1979).

28. Stone, D., S. Hannula and C.Y. Li, Proc. 35th Electronic Components Conf., 46-51 (1985).

29. S.S. Manson, "Thermal Stress and Low Cycle Fatigue", McGraw Hill, New York (1966).

30. Hagge, J.K., Proc. IEPS 1982, 198-208.

31. Felske, K.E. and J. Moss, Proc. 36th Electronic Components Conf., 641-8 (1986).

32. Kamei, T., M. Nakamura, H. Ariyoshi and M. Doken, Proc. 28th Electronic Components Conf. (1978).

33. Fox, L.R., J.W. Sofia and M.C. Shine, IEEE CHMT-4(2), 275-85 (1985).

34. Lauer, L.D., Circuit World 13(1), 13-7 (1986).

35. Frear, D., D. Grivas, M. McCormack, D. Tribula and J.W. Morris, Proc. ASM 3rd Conference on Electronic Packaging, 269-74 (1987).

36. Wright, E.A. and W.M. Wolverton, Proc. 34th Electronic Components Conf., 149-55 (1984).

37. Burgess, J.F., R.O. Carson, H.H. Glascock, C.A. Neugebauer and H.F. Webster, IEEE CHMT-7(4), 405-10 (1984).

38. W. Engelmaier, IEEE CHMT-6(3), 232-7 (1983).

39. Howard, R.T., S.W. Sobeck and C. Sanetra, Solid State Technology, Feb. 1983, 115-22.

40. Burgess, J.F., A.J. Yerman, J.F. Burgess and C.A. Neugebauer, Proc. 33rd Electronic Components Conf., 138-41 (1983).

41. Engelmaier, W., Proc. IEPS 1982, 15-23.

42. Olsen, D.R. and H.M. Berg, IEEE CHMT-2(2), 257-63 (1979).

43. Fisher, J.R., Proc. 34th Electronic Components Conf., 219-26 (1984).

44. Taylor, J.R. and D.J. Pedder, Proc. ISHM 1982 (209-14).

45. Wild, R.N., Proc. NEPCON 1974, 105-17.

46. Vaynman, S., M.E. Fine and D.A. Jeannotte, Met. Trans. 19A, 1051-9 (1988).

47. Thang Bui-Quoc, Testing and Evaluation 1(1), 58-64 (1973).

48. Solomon, H.D., Fatigue of 60/40 Solder, GE Technical Information Series, Class 1 (May 1986).

49. Ellozy, A.R., P.M. Dixon and R.N. Wild, Proc. Second Annual Conf. on Behavior of Materials, 903 (1976).

50. Coffin, L.F. Jr., Journal of Materials 6(2), 388-401 (1971).
51. Norris, K.C. and A.H. Landzberg, IBM Journal of Research and Development, 13(3), 266-71 (1969).
52. Manson, S.S., G.R. Halford and M.H. Hirschberg, Proc. ASME Symposium on Design for Elevated Temperature Environment, 12-28 (1971).
53. Shine, M.C., L.C. Fox and J.W. Sofia, Proc. IEPS 1985, 346-59.
54. Ostergren, W.J., Testing and Evaluation 4(5), 327-39 (1976).
55. Wilsching, P.H. and J.E. Kempert, Machine Design, July 8 1976, 65-9.
56. Sinclair, N.A., Proc. International Electrical Conference 1982, 56.
57. Solomon, H.D., Met. Trans. 4, 341-7 (1973).
58. Dance, F.J., Proc. Northcon 1983, 1-5.
59. Hanson, J.R. and J.L. Hauser, Electronic Packaging and Production, Nov. 1986, 45-51.
60. Greer, S.E., Proc. 28th Electronic Components Conf., 166-71 (1978).
61. Williams, R.R. and A.W. Noblett, Proc. First International SAMPE Electronics Conf., Vol. 1, 116-24 (1987).
62. Amick, P.J. and P.A. Cook, Proc. NEPCON West 1987, 661-7.
63. Ohshima, M., R. Satoh, K. Hirota and I. Ishi, Proc. First IEEE CHMT Symposium, August 1984, 165-70 (1985).
64. Fisher, J.R., Proc. 34th Electronic Components Conf., 219-26 (1984).
65. Cherian, G., Proc. IEPS 1984, 702-10.
66. Matsui, N., S. Sasaki and T. Ohsaki, Proc. 37th Electronic Components Conf., 573-8 (1987).
67. Koopman, N. and P. Totta, paper in Session VI of these proceedings.
68. Shukla, R.K. and N.P. Mencinger, Solid State Technology, July 1985, 67-74.
69. Winder, L. and R. Dietz, Proc. IEPS 1982, 715-27.
70. Moghadan, F.K., Proc. ISHM 1982, 79-87.
71. Tan, K.S. and D. Bose, Solid State Technology, April 1986, 165-8.
72. Hund, T.D. and S.N. Burchett, Proc. ISHM 1983, 243-50.
73. Van Kessel, C.G.M., S.A. Gee and J.J. Murphy, Proc. 33rd Electronic Components Conf., 237-44 (1983).
74. Riney, T.D., Applied Physics, 32(3), 454-60, (1961).
75. Goldmann, L.S., R.J. Herdzik, N.G. Koopman and V.C. Marcotte, IEEE PHP-13(3), 194-7 (1977).
76. Howard, R.T., IBM Journal of Research and Development, 26(3), 372-89 (1982).
77. Munford, J.W., IEEE PHP-11(4), 296-304 (1975).

WIRE BONDING AND TAB TECHNOLOGY

J. Loy
Honeywell
Plymouth, Minnesota, USA

INTRODUCTION

Wirebonding and Tape Automated Bonding (TAB) are two of the three methods used to provide electrical connection to semiconductor devices. The third method used is flip chip solder bump technology which has been combined with TAB technology to improve electrical performance of VLSI device interconnection.

The materials required for wirebond technology are gold and aluminum. The processes used are thermosonic and thermocompression bonding. The process descriptions and equipment capabilities will be discussed in this paper. The material variations are larger for TAB technology including both metals and organics. The electrically conductive material is copper which is often finished with gold or solder. The processes used in TAB assembly also vary widely. TAB processes include thermocompression, gold/tin eutectic, solder reflow and laser joining (welding). Examples of metallurgy resulting from these processes are shown in the art figures of this paper.

INTERCONNECTION OPTIONS

There are three prominent interconnection technologies that have been used to attach chips to packages. Flip chip, TAB and wirebonding have been used for over 20 years. These basic technologies have been extended and modified to accommodate today's high lead count chips. Today's VLSI (Very Large Scale Integrated Circuit) chips require 200 to 600 leads and also require sophisticated packages that can perform at frequencies greater than 100 megacycles. The selection of one option is based on the various requirements for the packaged integrated circuit in the system. These requirements are both mechanical and electrical.

Wirebond interconnection technology has been developed over the last 30 years. Today equipment is computer programmable with automatic operation modes. The materials used are gold and aluminum wires with 0.5-mil to 5.0-mil diameters. The process used for both gold and aluminum wire has evolved to a thermosonic bonding operation. The wirebonding equipment has improved in both speed and process uniformity. Special emphasis was placed on speed and uniformity as the leads per device increased rapidly during the 1980s. Control of the process variables is critical to achieving acceptable yields for high lead count devices. The process yield is more critical for multichip assemblies of high lead count devices. The development of computer-controlled process equipment has made it possible to extend the wirebond technology to VLSI Assembly Process. Examples of the progression of wirebond technology are shown in Figures 1, 2, and 3.

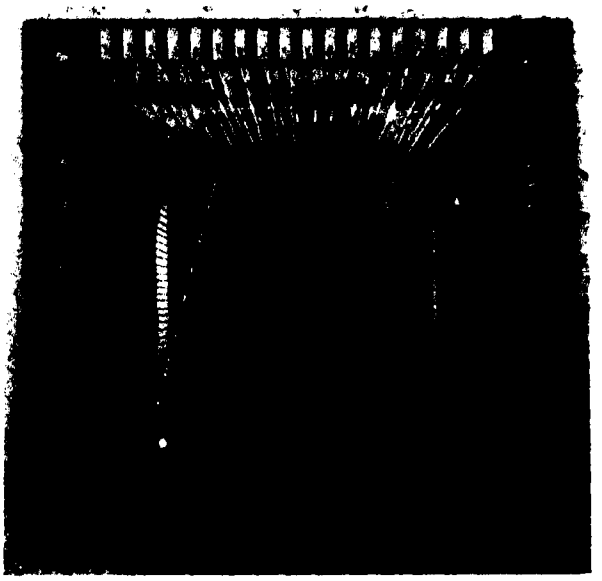


Figure 1. A two-tier multilayer ceramic pin-grid array package for a 144-lead wirebond device.

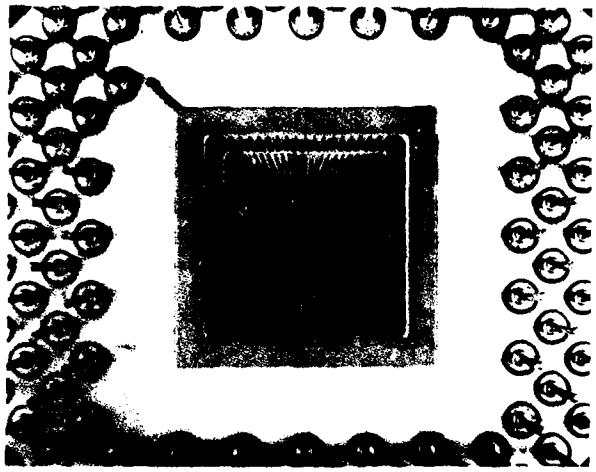


Figure 2. A two-tier multilayer ceramic for a 173-lead wirebond device.

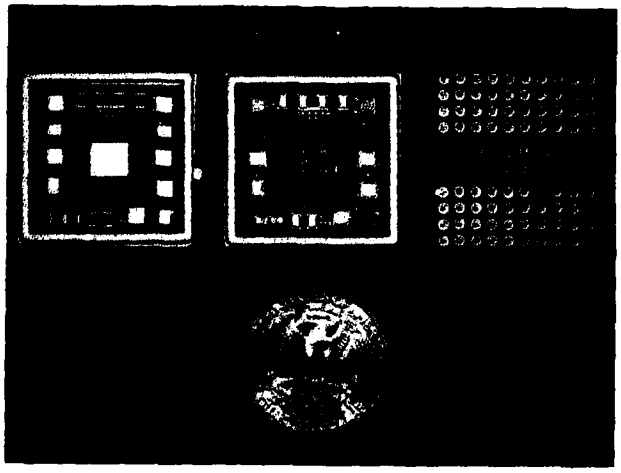


Figure 3. A multichip multilayer ceramic package utilizing wirebond interconnections.

TAB technology was developed as a potential cost-effective interconnection process for electrically attaching integrated circuits to packages. TAB technology has evolved since the mid 60s to its current state-of-the-art status of having two main applications. The first application is primarily based on low cost, high-yield assembly process for low to moderate complexity integrated circuits. The second application of TAB technology is that of electrical high performance interconnection processes for the most complex integrated circuits. TAB technology requires that bumps be placed on either the integrated circuit or the TAB. Both bumping techniques are being used and continue to be developed by the industry. Gold and solder are the materials predominately used for bumping wafers. Copper is the primary conductor material used for TAB fabrication and is used for adding bumps to TAB.

TAB assembly technology was developed using gold bumps on integrated circuits with a gold or tin plating finish added to the copper conductors of the TAB leadframe. The gold bump TAB assembly utilized the thermocompression bonding process, as shown in Figure 4. In a later development, Honeywell established a solder bump TAB assembly process. The solder TAB process combines solder bump flip technology with TAB for the high lead count integrated circuits.

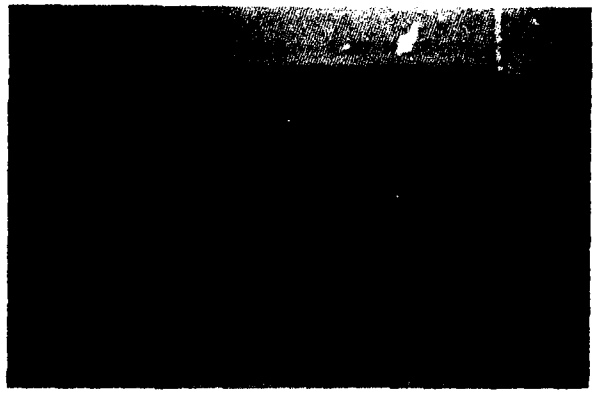


Figure 4. A gold bumped device assembled to TAB leadframe.

Solder bumps are processed by evaporation or plating techniques. An example of plated solder bumps is shown in Figure 5. Figure 6 shows a solder bumped device inner lead bonded to a TAB leadframe. The assembly process is reflow soldering.

Figure 5. SEM photo of plated solder bumps reflowed prior to TAB assembly.

COMPARISONS

Two comparisons of current packaging assembly with wirebonding and TAB are shown in Figures 7 and 8. Cost, yield, electrical performance, reliability, and system requirements must be considered in selecting an assembly process for VLSI devices.

Figure 6. SEM photo of a section of a solder bump TAB assembly (Inner-Lead Bond).

The factors of most importance when comparing wirebonding with TAB for VLSI devices, are pad density and electrical performance. Wirebonding density with gold wire is limited to 3.0 mil x 3.0 mil minimum size. TAB bonding allows a higher density of 2.0 mil x 2.0 mil minimum size. In addition, TAB bonding with solder bumps incorporate area pads including bumps over active areas of the silicon. The electrical performance comparison for high speed requirements indicates an advantage for TAB. Creative designs allow power and ground distribution to the integrated circuit with minimum resistance and inductance. In multichip modules, TAB offers the ability to test and burn-in devices prior to the final assembly. Both wirebonding and TAB can be manufactured to achieve required reliability levels. Both technologies require rigid process controls to maintain high assembly yields. The semiconductor equipment manufacturers are responding to the requirements of high lead count VLSI devices.

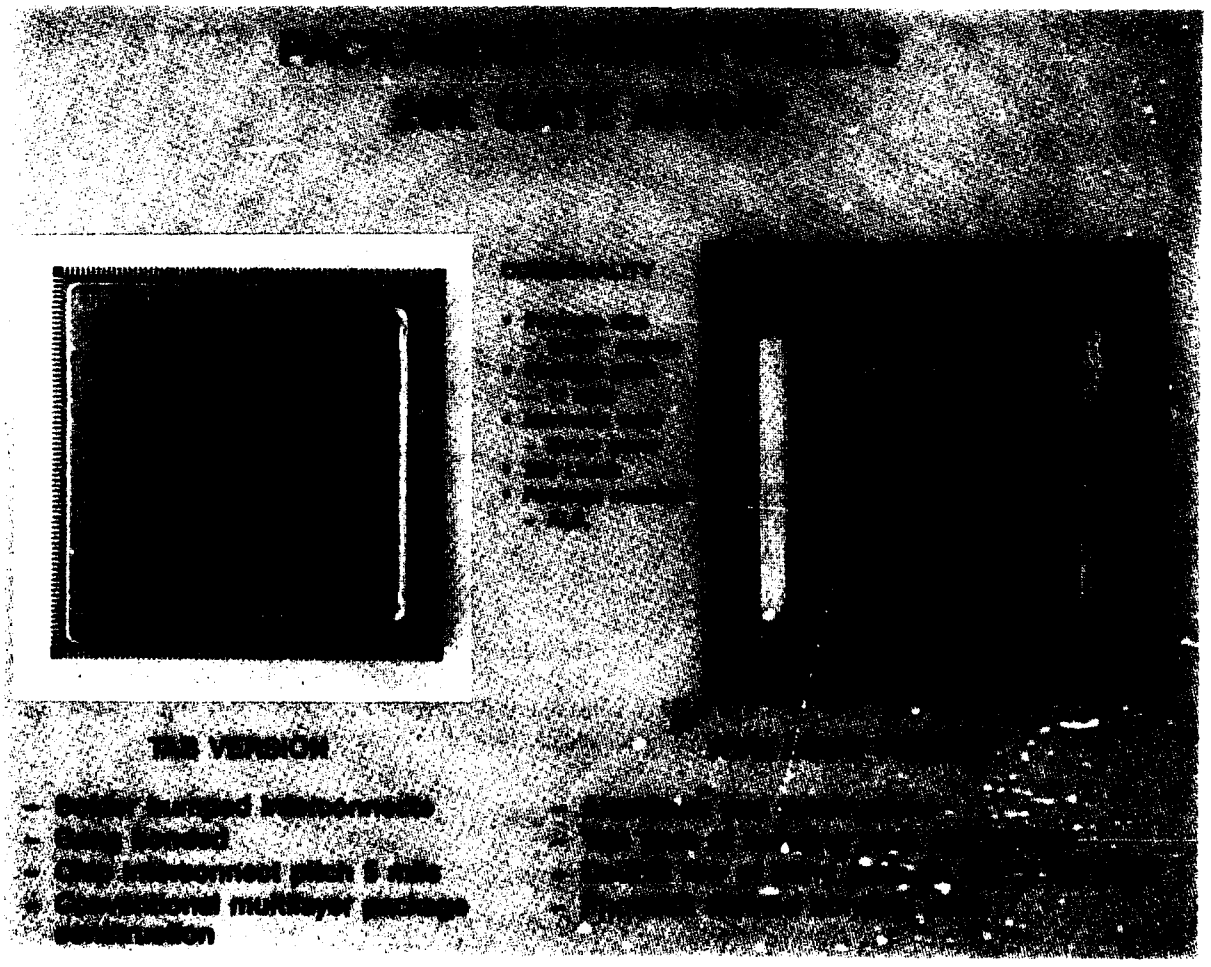


Figure 7. A comparison of TAB/wirebonded packages for a gate-array device.

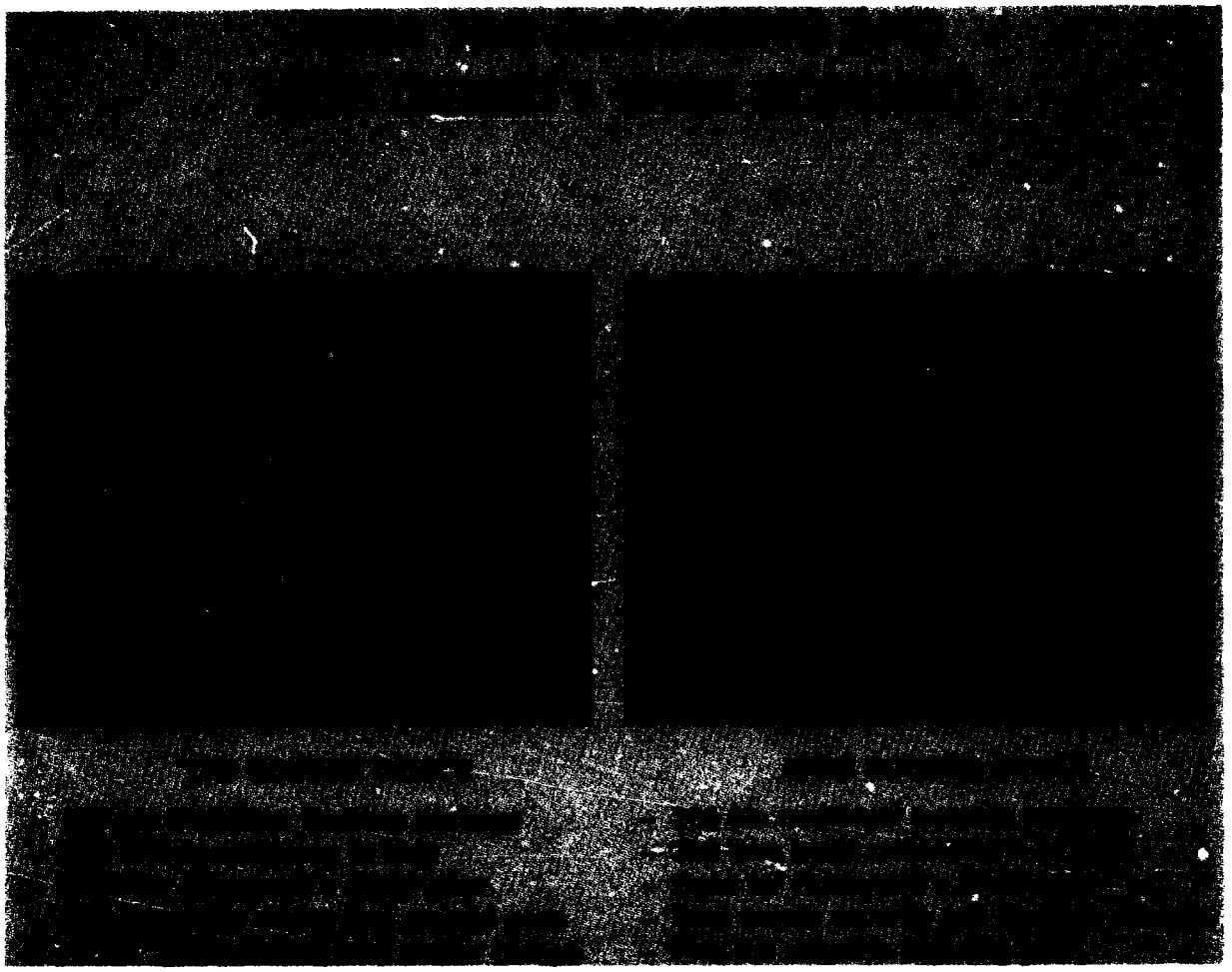


Figure 8. A comparison of TAB/wirebonded packages for two custom devices.

CONCLUSION

Wirebond technology and equipment suppliers have responded to the challenge of lead counts greater than 300 during the 1980s. The current TAB development activity is directed at VLSI devices with lead counts of 600 to 800. The industry trend indicates that lead counts for integrated circuits will reach 1000 in the early 1990s. The interconnection technology for the 1990s will require additional research and development.

CONTROLLED COLLAPSE CHIP CONNECTIONS (C4)

Nicholas G. Koopman, Paul A. Totta
IBM
East Fishkill, New York, USA

ABSTRACT

An in-depth review of Controlled Collapse Chip Connections (C4), otherwise known as Flip-Chip Technology, is presented. Materials, design factors, fabrication processes, tools, assembly, rework, and reliability are discussed. Future trends are indicated, and comparisons made with Wire Bonding and TAB (Tape Automated Bonding).

THE SOLDER-BUMP interconnection was initiated in the early 1960s to eliminate the expense, unreliability, and low productivity of manual wire bonding. Whereas the initial, low-complexity chips had typically peripheral contacts, this technology has allowed considerable extendibility in I/O density as it progressed to full population area arrays. The so-called Controlled Collapse Chip Connection (C^4 or C4) utilizes solder bumps deposited on wettable metal terminals on the chip and a matching footprint of solder wettable terminals on the substrate. The upside-down chip (flip chip) is aligned to the substrate, and all joints are made simultaneously by reflowing the solder (Fig. 1).

The C4 joining process has been described in the literature with numerous acronyms. C4, as originally used in IBM, denoted Controlled Collapse Chip Connection. In the industry, it has also been called CCB — Controlled Collapse Bonding— and Flip-Chip Joining, referring to the fact that it is opposite to the traditional back-side-down method of bonding, in which the active side of the chip, facing up, is wire bonded. The term C4 or flip chip is used here.

Two other acronyms are used in this section: BLM and TSM. These refer to the terminal connecting metallurgies at the chip and substrate, respectively. BLM stands for Ball Limiting Metallurgy and refers to the region of terminal metallurgy on the top surface of the chip that is wettable by the solder. TSM stands for Top Side Metallurgy and refers to the terminal metallurgy on the substrate to which the chip and its associated solder ball are joined.

C4 HISTORY — The solder-bump interconnection of flip chips, the face-down soldering of silicon devices to alumina substrates, has been practiced for approximately twenty years (3). First introduced in 1964 with the Solid Logic Technology (SLT) hybrid modules of IBM's System/360, it was part of a design to eliminate the expense, the unreliability, and the low productivity of manual wire bonding (2). The solder bump was also an integral part of a chip-level hermetic sealing system created by the glass passivation film on the wafer (4). Most semiconductor devices of that era were, in contrast, protected by expensive hermetically sealed cans. The terminal metallurgy design was intended to reseal the access of "via" holes through the glass as well as to provide a means for testing and joining the chip (Fig. 2).

Initially, for the discrete transistors or diodes of the hybrid SLT, copper ball standoffs, embedded in the solder bumps, were used to keep the unpassivated silicon edges of the chips from electrically shorting to solder-coated thick-film lands (5). Later, in the integrated circuit era, the controlled collapse chip connection was devised. In this technique a pure solder bump was restrained from collapsing or flowing

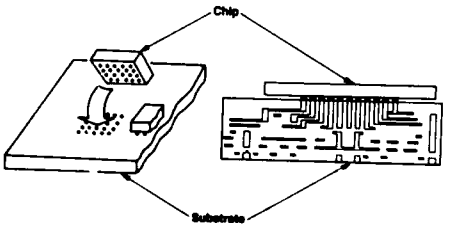


Fig. 1. Controlled Collapse Chip Connection (C4)
The upside-down chip (flip chip) is aligned to the substrate and all joints are made simultaneously by reflowing the solder. (After Ohshima, Ref. (1), 1982 and After Fried, Ref. (2), 1982).

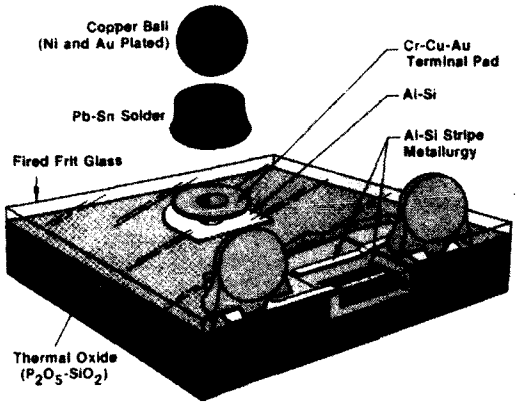


Fig. 2. Terminal Metallurgy Design
The original SLT flip chip (27 mm square) with glass passivation, BLM sealing of via holes, and *Cu* ball solder bumps. (After Totta, Ref. (5), 1969).

out on the electrode land by using thick-film glass dams, or stop-offs (6), which limited solder flow to the tip of the substrate metallization (Fig. 3).

Similarly, the flow on the chip is limited by a ball limiting metallurgy (BLM) pad, which is a circular pad of evaporated, thin-film chromium, copper, and gold that provides the sealing of the via as well as a solderable, conductive base for the solder bump. A very thick deposit (100–125 μm) of evaporated 95 *Pb*/5 *Sn* solder acts as the primary conduction and joining material between chip and substrate (7).

The early integrated circuit chips typically had peripheral C4 I/O pads like their wire-bonded counterparts. The pads were 125 to 150 μm in diameter, on 300 to 375 μm centers. The pitch of connections was typically compatible with the resolution capability of thick-film (*AgPd*) electrode screening on the ceramic substrates.

Occasionally it was convenient to have an "in-board" power pad or two in the thick-film technology, but larger numbers of inside I/O pads could not be used until metallized ceramic (MC) technology became available in the mid-1970s. The narrower lines and spaces made possible with etched thin-film *Cr–Cu–Cr* on ceramic, associated with that technology, allowed wiring escape for double rows of I/O pads and many internal connections. Later, a "depopulated" grid of bumps allowed the interconnection of 120 pads on 700 circuit logic chips. The fully populated area grid array, in which every grid point is occupied by a solder bump, required the complexity of multilayered, cofired ceramic packages. In these packages the distribution of I/O wiring could be accommodated by via "microsockets" and buried layers of wiring as opposed to single-level wiring where the "escape" of wires is geometrically restricted by the maximum number of lines per channel between I/O connections (8). A C4 structure with multilayer ceramic is shown in Fig. 4.

An example of an area array C4 configuration is shown in Fig. 5. The I/O count is 120 in an efficient square grid array, which is 11 C4 pads long by 11 pads wide on 250μm (10 mil) centers. A 125μm (5mil) solder bump is located at every intersection in the grid except one, which is displaced for orientation purposes (2). The matching multilayer ceramic substrate is shown in Fig. 6. Some of these co-fired alumina ceramic packages use 9 to 133 area array chip sites per package to attain high bipolar circuit densities in IBM's 4300 and 3081 series computers. Logic and memory chips are mixed as required. As many as 25,000 logic circuits or 300,000 memory bits have been packaged on a single TCM

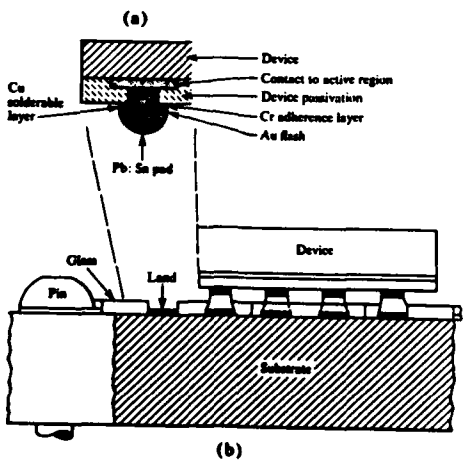


Fig. 3. Controlled Collapse Chip Connection
(a) Side view of a device. (b) Side view of a device on a substrate (dam method). (After Miller, Ref. (6), 1969).

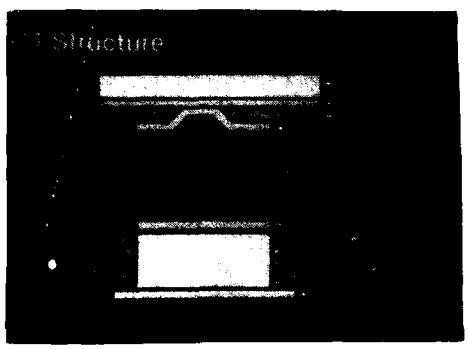


Fig. 4. C4 Structure
Schematic of a C4 joint on a multilayer ceramic "microsocket".

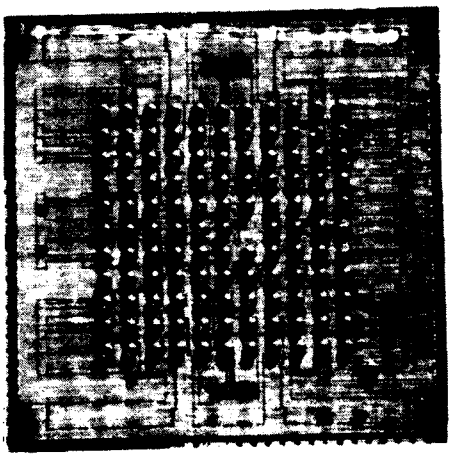


Fig. 5. Area Array C4 Configuration
An 11 x 11 full area array of solder bumps on a 700-circuit logic chip for use with multilayered ceramic. (After Goldmann, Ref. (3), 1983, reprinted with permission of Solid State Technology).

substrate in this technology (8), (9), (10), (11), (12). The most populous logic flip-chip to date is a "computer-on-a-chip" which has 762 solder bumps in a 29 x 29 area array Fig. 7. This chip, with four levels of metal wiring, is used in IBMs 9370 computer (13) — two or three levels of metal wiring were used in earlier systems.

Over time, the C4 technology has been extended to other applications. C4s are used on thin-film resistor and capacitor chips in hybrid module applications (14). Solder pads for this application are very large—750 μm in diameter.

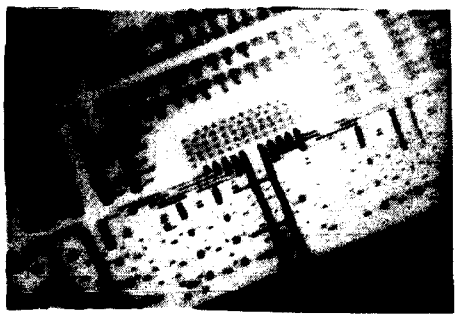


Fig. 6. Redistribution Connections Near Top of MLC
Section through area array chip site of a multilayered ceramic substrate showing redistribution connections near top of MLC. (After Goldmann, Ref. (3), 1983, reprinted with permission of Solid State Technology).

At the other extreme, Schmid (15) used C4s for precision registration and alignment in the joining of a *GaAs* waveguide. The C4s in this case were only 25 μm high. The most dense area array reported has been a 128 x 128 array of 25 μm bumps on 60 μm centers, resulting in 16,000 pads (16). The photolith process for forming these is discussed later. C4s or solder structures similar to C4s are used for attaching chip carriers to boards and have become part of the surface mount revolution discussed in the first chapter. New applications of this technology are being explored continually.

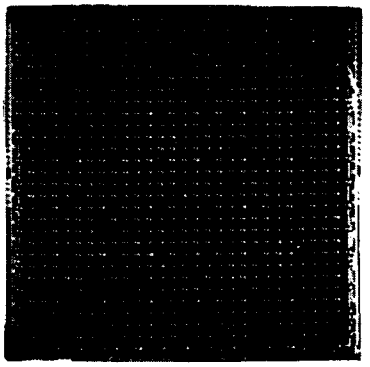


Fig. 7. Computer-on-a-Chip
An 11,000-circuit "Computer-on-a-Chip" with 762 solder bumps in 29 x 29 area array. Courtesy of IBM Corp.

MATERIALS — Melting point has been a prime consideration in the choice of solder alloys for C4s. High-lead solders, especially 95 *Pb*/5*Sn*, have been most widely used with alumina ceramic substrates because of their high melting point, approximately 315°C. Their use for the chip connection allows other, lower melting point solders to be used at the module-to-card or card-to-board packaging level without remelting the chip C4s.

A reverse order of assembly (e.g., modules-to-board, then chips-to-module) would require a reverse order of melting point. Josephson superconducting devices have been joined in such a fashion, with an alloy of 51 *In*/32.5 *Bi*/16.5 *Sn* (having a melting point of 60°C), being used for the chip C4s, while a higher melting point eutectic alloy, 52 *In*/48*Sn* (having a melting point of 117°C), was used for pins and for orthogonal connections to the chip carrier (17), (18), (19).

Joining to organic carriers such as Kevlar-polyimide (20) or printed-circuit boards (21), (22), (23) also requires lower processing temperatures. Here, intermediate melting point solders such as eutectic *Sn* 63/*Pb* 37 (melting point 183°C and *PbIn* alloys such as *Pb* 50/*In* 50, with a melting point of approximately 220°C) have been used. A listing of solder alloy compositions and melting points is shown in Fig. 8.

Some phase diagrams relevant to C4 solder joints are shown in Fig. 9.

The choice of terminal metals, which is described in detail later, will likely depend upon the choice of solder. For example, silver and gold are poor terminal metals to be used with the *SnPb* alloy. In only a few seconds gold completely dissolves into the liquid solder. In these cases another solder alloy could be used, such as Indium (15), which has a much lower solubility for gold; or one of the other lower-solubility metals could be used for the terminal. Thus, *Cu*, *Pd*, *Pt*, and *Ni* are very commonly used for both BLM and TSM thin films. All of these metals form intermetallics with *Sn*, which limits the reaction rates with *PbSn* solders. On the chip side, this terminal metal is normally sandwiched between an adhesion metal layer of *Cr* or *Ti*, and a passivation metal layer, usually of thin gold. The copper, palladium, or nickel thin films on the substrate are either similarly protected with gold (9), (10), (11), (12) or are dip-soldered. In the latter case, some of the solder for the C4 joint is supplied by the substrate (25), (26). MLC substrates usually use a flash of gold on nickel (1), (9), (10), (11), (12), (14), (27), (28). Thick-film substrates have the palladium or platinum alloyed with gold or silver and are dip-soldered prior to the joining operation. *AuPt*, *AgPd*, *AgPdAu*, and *AgPt* have been reported (6), (7), (29), (30), (31), (32), (33) as thick-film TSM pads.

DESIGN FACTORS — Some of the factors affecting the material choices for the terminal and solder have already been discussed, but other variables that must also be considered in C4 design. The joints must be high enough to compensate for substrate nonplanarity, especially for the older version of thick-film substrates. Because solder surface tension "holds up the chip," a sufficient numbers of pads must be provided to support the weight of the chip. Typically this becomes a cause for concern only with very low I/O devices such as memory chips or chip carriers, which are very bulky. Numerous studies have been published showing the interrelationship between BLM and TSM size, solder volume, chip weight, and C4 height. Fig. 10 shows these relationships over a very wide range (14).

Extra "dummy" pads, added to supplement those needed for simple electrical connection, have often been used to enhance the mechanical behavior, reliability, or thermal performance of the assembly (34). Pad location is important from the standpoints of both electrical design and reliability. The effect of distance to neutral point (DNP) is discussed later as it relates to thermal cycle fatigue.

As VLSI chips become more and more dense, higher I/O counts will drive full area arrays of terminals. In this case, the pad size and location are fixed by the chip size and available real estate per pad allocated by a fully populated area array.

The number of C4 pads as a function of chip size and pad geometries is shown in Fig. 11 wherein the possibility of 155,000 pads on a 20 mm chip is indicated. Fig. 12 shows the pronounced density advantage of area array versus a single-perimeter row, as pad sizes and spacings decrease.

Fig. 8. Selection of Low Melting Solder Alloys
(After Wassink, Ref. (24), 1984, reprinted with permission of Electrochemical Publications Ltd., Ayr, Scotland).

Melting Point	Composition in Mass Percent					
°C	*Sn*	*Pb*	*Bi*	*In*	*Cd*	Other Elements
16				24		76 *Ga*
20	8					92 *Ga*
25						95 *Ga*; 5 *Zn*
29.8						100*Ga*
46.5	10.8	22.4	40.6	18	8.2	
47.2	8.3	22.6	44.7	19.1	5.3	
58	12	18	49	21		
61	16		33	51		
70	13.1	27.3	49.5		10.1	
70–74	12.5	25	50		12.5	
72.4			34	66		
79	17		57	26		
91.5		40.2	51.7		8.1	
93	42			44	14	
95	18.7	31.3	50			
96	16	32	52			
96–98	25	25	50			
103.0	26		53.5		20.5	
96–110	22	28	50			
117	48			52		
125		43.5	56.5			
127.7				75	25	
139	43		57			
144			62		38	
145	49.8	32			18.2	
156.4				100		
170	57				43 Tl	
176	67				33	
178	62.5	36				1.5 *Ag*
180	63	34	3			
183	61.9	38.1				
183	62	38				
198	91					9 *Zn*
215		85				15 *Au*
221	96.5					3.5 *Ag*
232	100					
248		82.6			17.4	
251		89				11 *Sb*
266					82.6	17.4 *Zn*
271			100			
280	20					80 *Au*
288		97.2				2.8 *As*
304		97.5				2.5 *Ag*
304–312	5	95				
318		99.5				0.5 *Zn*
321					100	
327		100				
356						88 *Au*; 12 *Ge*
370						97 *Au*; 3 *Si*
420						100 *Zn*

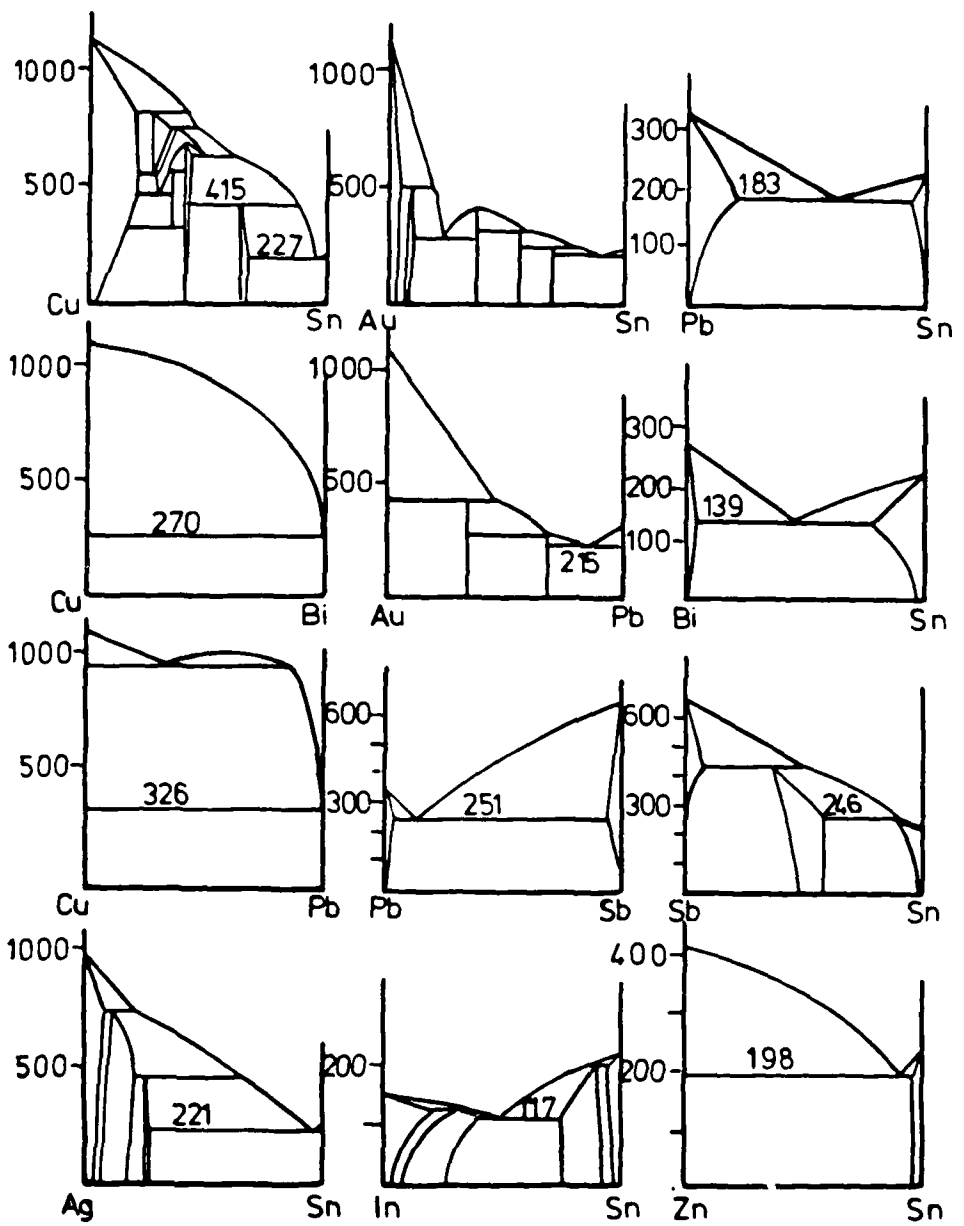


Fig. 9. Some Phase Diagrams Relevant to Soldering.
Temperature, along the vertical axes, is given in °C; Concentration, along the horizontal axes, is given in mass percent (Hansen and Smithells give concentrations in atomic percentages). (After Wassink, Ref. (24), 1984, reprinted with permission of Electrochemical Publications Ltd., Ayr, Scotland).

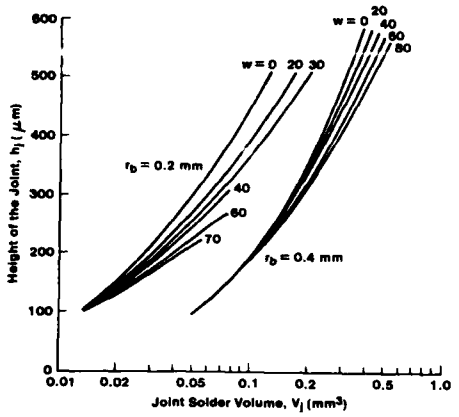


Fig. 10. Interrelationship between BLM and TSM Size Height of the joint (hj) vs. solder volume of the joint (Vj) as a function of chip weight (W in grams); r_b is radius of joint. (After Kamei, Ref. (14), 1978, ©IEEE).

FABRICATION PROCESSES AND TOOLS — Metal mask technology (3), (5), (7), (14), (18), (30), (31), (36), (37) is most widely used for terminal fabrication. BLM and solder are evaporated through holes in a metal mask and deposited as an array of pads onto the wafer surface. This is a low-cost batch process, simultaneously involving many chips per wafer and many wafers per evaporation. DC sputter cleaning of the via hole is commonly used to remove undesirable oxides and photoresist just prior to evaporation (5). This step assures low contact resistance to the aluminum metallurgy of the device and good adhesion to the SiO_2 or polyimide insulator.

The multilayered structure of the BLM can be described by using *Cr–Cu–Au* as an example. A typical evaporator would have numerous metal charges with thermal energy supplied by resistance, induction, or electron beams (e-Guns). *Cr* is evaporated first to provide adhesion to the passivation layer, as well as to form a solder reaction barrier to the aluminum. A phased layer of *Cr* and *Cu* are coevaporated next, to provide resistance to multiple reflows. This is followed by a pure *Cu* layer, to form the solderable metallurgy. A flash of gold is then provided as an oxidation protection layer. This is necessary because the wafers are normally exposed to air before going on to the next step of solder evaporation. This is performed through the same mask as the BLM but in a separate evaporator. This requires a "thick" (of the order of 100μ thick) mask for the solder. While lead and tin are usually in the same charge (single molten pool), the higher vapor pressure component, *Pb*, deposits first, followed by tin on top of the lead. Reflow in a H_2 ambient furnace at about 350°C melts and homogenizes the pad and brings it to a spherical shape. In addition H_2 assures reduction of oxides of *Pb* and *Sn*.

Fig. 11. Pad Density Comparisons (After Nye, Ref. (35), 1986).

Chip Size mm	Pad Dia. μm	Pad Pitch μm	No. of Pads Perimeter	No. of Pads Array
	762	1524	8	9
	635	1270	12	16
	508	1016	16	25
	381	762	24	49
	305	610	28	64
5	254	508	36	100
	203	406	44	144
	152	305	60	256
	127	254	76	400
	102	203	96	625
	76	152	128	1,089
	51	102	192	2,401
	25	51	388	9,604
	762	1524	21	49
	635	1270	28	64
	508	1016	36	100
	381	762	48	169
	305	610	60	256
10	254	508	76	400
	203	406	96	625
	152	305	128	1,089
	127	254	152	1,521
	102	203	192	2,401
	76	152	260	4,356
	51	102	388	9,604
	25	51	784	38,809
	762	1524	48	169
	635	1270	60	256
	508	1016	76	400
	381	762	100	676
	305	610	128	1,089
20	254	508	152	1,521
	203	406	192	2,401
	152	305	260	4,356
	127	254	312	6,241
	102	203	388	9,604
	76	152	520	17,161
	51	102	784	38,809
	25	51	1,572	155,236

Photolithographic processes and combinations of photolith and metal mask are becoming more and more popular for fabricating terminals (16), (27), (38), (39), (40), (41), (42), (43), (44), (45). Most common is a sequence of blanket deposition of the BLM, application of photoresist, development of a pattern in the resist followed by electrodeposition of the solder; then removal of the resist and subetching of the BLM, using the plated solder bumps as a "mask" (16), (38), (40), (42), (44). An alternative

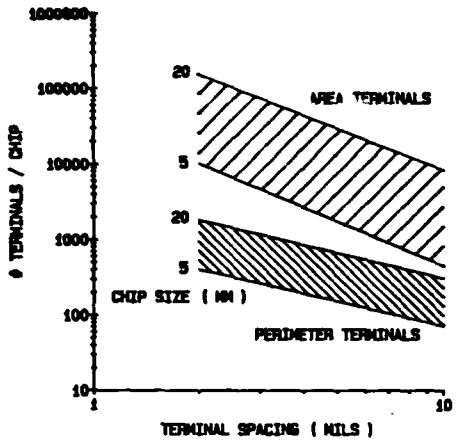


Fig. 12. Number of Pads vs. Pad Separation for Different Chip Sizes
Input/output terminal trends. (After Nye, Ref. (35), 1986).

sequence is to blanket-deposit BLM, photoresist, subetch BLM through the resist, then deposit solder by a variety of techniques including solder dip (39), solder ball placement (45), or metal mask evaporation (41). The single masking processes (unimask) are significantly simpler and cheaper than multiple mask techniques, although they are not as flexible in providing varying amounts of solder. In some applications higher volumes of solder are able to decrease strain.

The final operation before dicing of the wafer into individual chips is the electrical testing of each chip. Mechanical probes are used to contact the soft solder bumps fabricated earlier.

The formation of wettable surface contacts on the substrate (providing a mirror image to the chip contacts) is achieved by thick or thin-film technologies. Thin-film contact technology is similar to the BLM described previously, but thick-film technology involves development of wettable surfaces by plating nickel and gold over generally nonwettable surfaces such as *Mo* or *W* (conductors usually used within the ceramic substrate). Solder flow is restricted by the use of glass or chromium dams where necessary.

Various thin and thick-film processes that are typically used are shown in Fig. 13.

ASSEMBLY/REWORK — Once the BLM, TSM, and solder are in place, as described previously, the joining of chips to the substrate using C4 technology is straightforward. Flux, either water-white rosin (42), (46), (47), (48) for high-lead solders or water-soluble flux (17), (18), (49) for low-lead and other low-melting solders, is normally placed on the substrate as a temporary adhesive to hold the chips in place. Such an assembly is then subjected to a reflow thermal cycle involving either individual chip joining using a local heat source to bond one chip at a time, (20) or joining an assembly of several chips to a substrate simultaneously using an oven (47).

One of the greatest features of the C4 process is its self alignment capability arising from the high surface tension forces of solders (37). Chip pads and their counterparts on the substrate may be separated by as much as three times the average bump radius, but if the mating surfaces touch and are reasonably wettable, self alignment will occur. As many as a million C4 bonds can be made in one hour using automated tools.

Once the chip-joining operation is complete, cleaning of flux residues is accomplished with such solvents as chlorinated solvents or xylene for rosin flux and water for water soluble flux. The assembly is then electrically tested.

Rework — Reworking may be necessary to replace defective chips on multichip modules or for engineering changes. In this case, one or more chips may be removed from the module and replaced with new ones. This process is described by Puttlitz (47) for large multilayer ceramic modules.

Device debonding can be achieved by mechanical means (torque or pull) or by melting the solder and directly lifting the chip from the surface with a vacuum pencil. Residual solder is left on the substrate, which must be removed prior

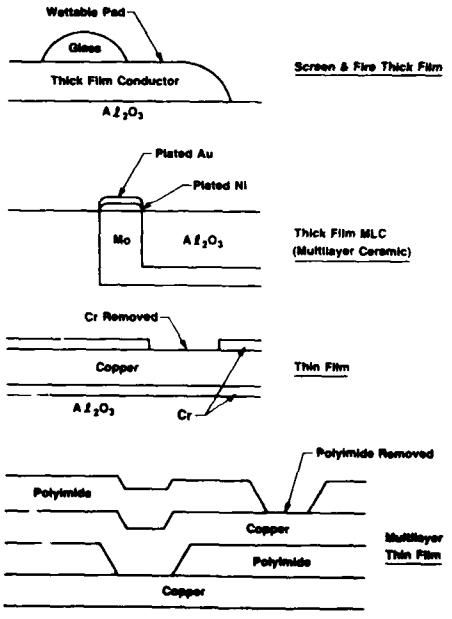


Fig. 13. Various Substrate Pad Structures (TSM)

to joining the new chips. The excess solder is removed from the substrate pads by means of a hot gas tool (Fig. 14), which ensures that adjacent sites are not contaminated with the excess solder. The TSM microsockets are then ready for chip joining. Multiple chip replacements per chip site have been performed using this method.

C4 Reliability — A question often raised regarding flip-chip bonding is the ability of the joint to maintain structural integrity and electrical continuity over a lifetime of module thermal cycling. A thermal expansivity mismatch between chip and substrate will cause a shear displacement to be applied on each terminal. Over the lifetime of a module, this may lead to an accumulated plastic deformation exceeding 1,000% (50). A quasi-empirical model was developed by Norris and Landzberg (29) that relates the cyclic lifetime to cyclic deformation parameters. It is based on the Coffin-Manson relationship (51) between fatigue life and plastic strain amplitude but with two terms added to account for time-dependent behavior: a frequency term, where lifetime increases with frequency to a low power, and a maximum temperature term, where lifetime decreases with maximum temperature. With the assumption of a log-normal failure distribution, a product sample may be tested in an accelerated thermal cycle; then, based on the statistical lifetime to electrical failure, the projected field lifetime may be extrapolated. Using this technique, an interconnection failure rate projection was made for logic chips in System/370 (29) of no more than 10^{-7}%/1,000 hours per bond at the end of life. By the end of 1975, 540 billion MST interconnection-hours had been accumulated with no wearout failures reported (52), yielding a 50% confidence level estimated failure-rate of 1.3 x 10^{-7}% /1,000 hours, in good agreement with the earlier projection.

These results demonstrate the inherent capability of solder interconnections to withstand high strain accumulations; the results also demonstrate the approximate validity of the projected failure rate. However, the early MST chip is only about 1 x 1 mm in size and has only 12 peripheral bumps. The mismatch shear deformation is proportional to the distance between a given pad and the neutral point referred to as DNP, (the point on the chip that remains stationary relative to the substrate during a thermal excursion). Because the neutral point is near the chip center, the maximum value of DNP is roughly proportional to chip size. Moreover, the Norris-Landzberg model and most subsequent experiments show that lifetime is inversely proportional to shear deformation, and thus DNP, raised to a power that approaches two. Thus, with the evolution from MST to much larger and denser C4 footprints, thermal wearout was considered with renewed interest.

The wearout model subsequently has come under close scrutiny. It has been suggested, for instance, that a dwell time factor be added (53). Also recommended is a complete

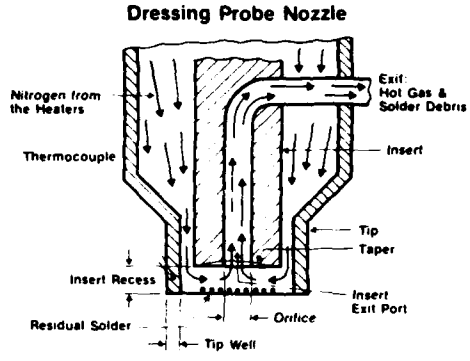


Fig. 14. Hot Gas Tool
Side-view sketch showing the direction of gas flow within the probe tip of the hot-gas dress tool. (After Puttlitz, Ref. (47), 1984, reprinted with permission from Journal of Electronic Materials).

reformulation starting with the constitutive equation for solder that incorporates crack growth and creep deformation (54). It has been further suggested that several competing mechanisms come into play in thermal wearout, including cavitation. Recent experimentation (55) also suggests that wearout is much more complex than implied by the simple equations discussed above. Scanning electron microscopy photos of joints at various stages of thermal cycling show that mismatch between the solder and chip plays a role in the damage. Also, low and high volumes of solder joints fail by different mechanisms, thus emphasizing the effect of C4 shape.

A further complicating factor in modeling wearout is that thermally induced strains are not uniform within the joint. An initial attempt (56) was made to incorporate joint shape into failure-rate projections. Despite the simplicity of the model, reasonable agreement has been obtained in experiments where joining geometry has been intentionally varied and the joints mechanically tested by single-cycle or cyclic torquing of the chip (14), (56). More sophisticated techniques, taking into account time-and temperature-dependent solder properties, will be required to understand geometric effects fully. Chip bending also must be included in a complete analysis,. (43), (57).

In summary, simple models that accompanied the introduction of C4 joining have, until now, proven adequate to estimate field behavior and as an aid in product design. This can be attributed not only to their simplicity and their qualitative rational but also primarily to the fact that existing products have experienced wearout failure rates too low to be of concern. As chips grow larger and pad counts in the hundreds become common, new or revised models will be required that reflect a greater understanding of the wearout mechanism and better precision in failure projection.

THERMAL MISMATCH RELIABILITY: EXTENSIONS — Together with a greater understanding must come an extension of the C4 in its ability to accommodate larger and denser chips without affecting system reliability. Existing modeling and testing techniques, albeit imperfect, have been used to evaluate various extension schemes, several of which show promise. They may be subdivided into strain reduction, geometry or shape improvement, and alternate solders.

The thermal mismatch displacement across the pad can be kept to a minimum by arranging the footprint to minimize the DNP, for example, by deleting corner pads or, in an extreme case, by a quasi-circular array (3).

Joint geometry is dictated by the wetting areas on chip and substrate, solder volume, and the weight of the chip. Unless the chip is very heavy, the joint has the shape of a spherical segment (14), (56), (61), and its height is uniquely determined by interface radii and volume. The previously mentioned geometric model (56) claims that wearout depends upon shape and that a joint can be geometrically optimized to extend lifetime. Normally, optimization of the spherical segment joint is of limited extendibility value, improving lifetime by less than 50%. Enforced changes in shape away from a spherical segment, on the other hand, can produce very large effects. Heavy chips that de ress the joint (14) severely reduce lifetime, while stretched or elongated joints substantially extend lifetime. Mechanical testing has shown an order of magnitude difference in fatigue life between an hourglass and a barrel-shaped joint, with the fracture location of the hourglass joint shifted to the center of the joint. (62), (58) Stretched pads have been fabricated by a number of techniques including using two different solders on the same chip (58). Called SST (Self-Stretching Soldering Technology), this technique makes use of the surface tension forces of larger bumps of one solder to stretch the lower volume functional bumps (Fig. 15). Two additional concepts being pursued in Japan are illustrated in Fig. 16 wherein solder columns are being stacked to achieve improved fatigue life.

In addition, solder has been cast into helical copper coils to form very high aspect ratio solder columns. It has been applied to joining leadless ceramic chip carriers (LCCC) to glass–epoxy printed circuit boards (Fig. 17). This structure has not been scaled down to sufficiently fine dimensions to be applicable to integrated circuit chip interconnections. Freestanding cast-solder pillars (without the copper helix) have been developed for package-to-board interconnections (23).

Among solders, which have been evaluated as alternatives to 95 *Pb*/5 *Sn*, the *Pb*/*In* system has shown the most fatigue enhancement. Thermal cycle lifetime has been shown to be quite sensitive to composition with a minimum at 15 to 20% *In*(46) , (65). There is a two times improvement over 95 *Pb*/5 *Sn* at 5% *In*, three times at 50%, and twenty times at 100% *In*(65) . Pure In is being used for optoelectronic device joining (15), (16). Early work on integrated circuits emphasized 50 *Pb*/50 *In* as a compromise between ultrahigh thermalcycle reliability and processing constraints. Implementation of this alloy was limited by two factors: increased corrosion susceptibility in nonhermetic packages (46), (65), (66), and a substantially accelerated thermomigration rate over *Pb*/*Sn* alloys (67). By the latter mechanism, the thermal gradient between chip and substrate causes a condensation of vacancies at the chip BLM region, leading to premature high resistance or mechanical wearout. Later work emphasized low (3–5%) In alloys (46), which provided less fatigue enhancement but were not susceptible to the corrosion and thermomigration difficulties of 50 *Pb*/50 *In*. Testing 95 *Pb*/5 *In* at several thermal cycle frequencies show similar relationship to that found by Norris and Landzberg for 95 *Pb*/5 *Sn* .

By far, the largest effort to reduce the strain drastically is by matching the substrate thermal expansion to that of silicon, as demonstrated in Fig. 18. This initial study using a Kevlar-polyimide organic substrate (20) has been followed by a number of packaging efforts that take advantage of the improvement in lifetime of the C4 connection as a result of thermal expansion match. The use of *AlN*, *SiC*, *Si*, and glass-ceramics as first-level packages and Kevlar-polyimide, copper-invar-copper boards with polyimide and copper thin-film wiring, are being actively pursued for direct chip-attach applications. Gallium arsenide has also been matched using sapphire (15). Power cycling is replacing thermal cycling to evaluate these material combinations (57), (68), (69). Process constraints, wirability, dielectric constant, and heat dissipation must, of course, be among other factors considered in selecting an alternative substrate material. Epoxy resins with coefficients of thermal expansion similar to that of solder ($260 \times 10^{-7}/°C$) are also being pursued to improve thermal fatigue life as shown in Fig. 19.

Among other reliability concerns that have been reported in the literature, thermomigration has been shown to be a major concern for applications where high-temperature gradients are coupled with high-diffusivity, low melting point

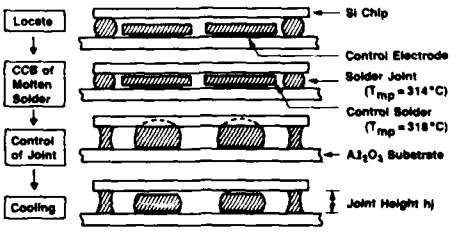


Fig. 15. Self-Stretching Soldering Technology (SST) (After Satoh, Ref. (58), 1983).

solders (67). Corrosion has been encountered in high-humidity testing of *PbIn*(66) . Solder void defects are addressed in (52). Palladium depletion of *AgPd* thick films by *PbSn* prompted a switch to *AgPdAu* ternary alloy for the substrate electrode (70), (71).

A more fundamental problem receiving increased attention is that of soft errors in devices caused by alpha particle emission of trace quantities of radioactive materials in the packaged assembly (26), (72), (73). In the case of C4, high-lead alloys are mostly used for computer applications, and lead almost invariably brings with it trace amounts of uranium and thorium. This problem will demand more attention as VLSI devices become more dense and as the critical charge levels in the device become smaller.

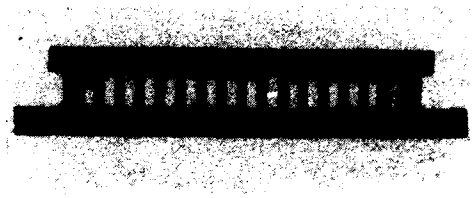


Fig. 17. Solder Columns
Close-up view of the solder columns connecting a LCCC to a glass–epoxy printed-circuit board. (After Cherian, Ref. (63), 1984).

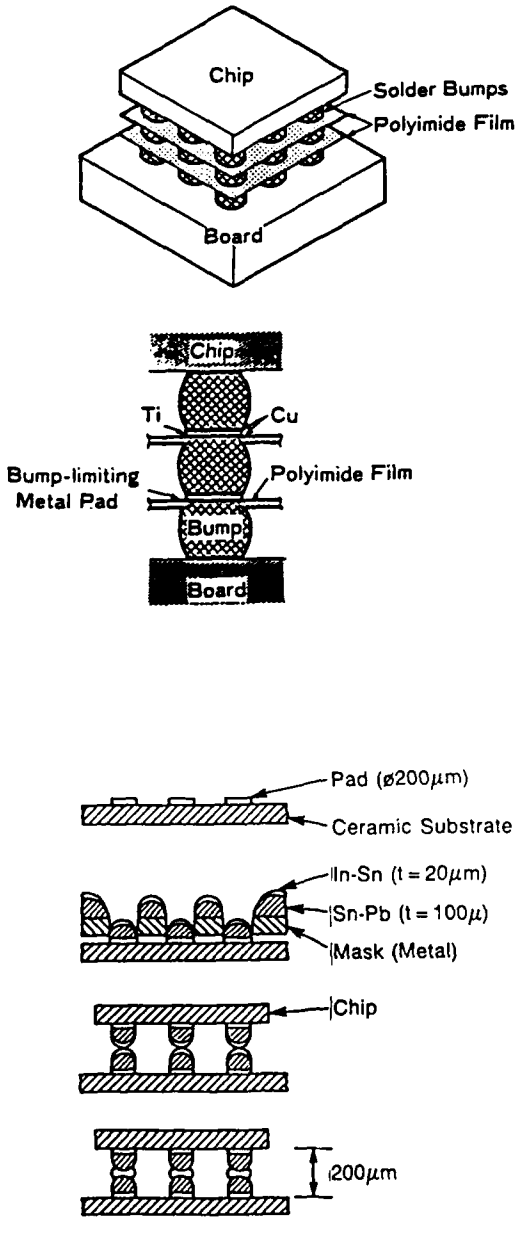


Fig. 16. Some Methods Being Pursued to Extend C4 Life
(a) stacked solder bumps using polyimide (After Matsui, (59), 1987 ©IEEE (b) stacked solder bumps using multiple solders (After Fujitsu (60)).

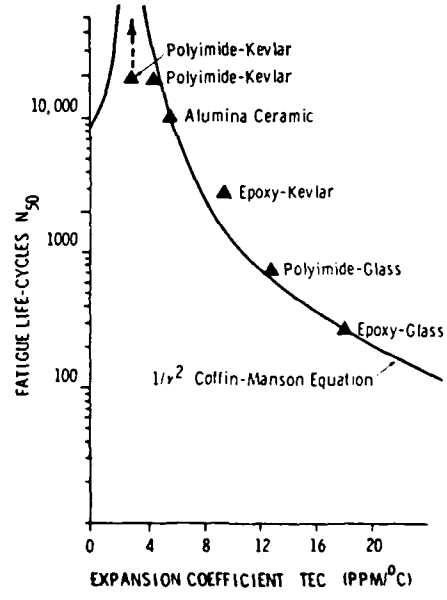


Fig. 18. Effect of Thermal Expansion Coefficient on Fatigue Life (After Greer, Ref. (20), 1978, ©IEEE)

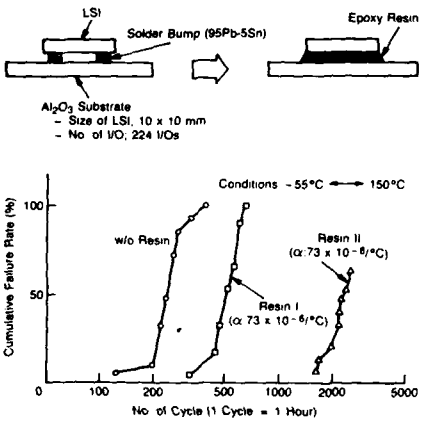


Fig. 19. C4 Life Extension by the use of Thermal Expansion Matched (to Solder) Resins
(After Nakano, (64), 1987)

HEAT DISSIPATION WITH C4 — A traditional constraint of all flip-chip bonding has been the poor heat dissipation capability compared to back-side bonding. Early chips with 10 to 20 peripheral C4 bumps bonded to an alumina substrate could dissipate approximately 0.5 Watt (74). While the interconnection portion of the heat transfer path was at least an order of magnitude more resistive than a comparable back bond (7), (75) a comparison of overall thermal resistance showed only about a 50% increase for flip chips for a typical module (75), (76). Thermal patterns for flip chips depend upon device location, size, metallization, and number of the terminals, and upon the thermal resistance of the substrate. Thus, numerical analysis techniques are required for reasonable performance projections (4), (74).

Area array bumps, in addition to meeting the density needs of VLSI, also provide enhanced dissipation because of the increased number of joints (effectively approaching a continuous heat transfer layer) and their greater proximity to the devices. Flip chips are becoming competitive thermally with back bonding. For instance, an air-cooled module with six 4.5 mm chips, each having a 11 x 11, 0.25 mm grid array of solder joints, can dissipate approximately 1.5 Watts per chip (8). Higher power levels have been achieved with new, high thermal conductivity ceramics such as *AlN*(77) and *SiC*(78).

Flip chips may have another thermal advantage compared to back bonding. Because the back of the chip is free of mechanically or electrically delicate surface features, it is thus amenable to direct contact by a wide variety of heat sinks whose thermal conductivity is often better than the plastic or ceramic package they are backbonded to. An example is the IBM multichip module (4), (79), where spring-loaded pistons transfer heat from the back of each chip to a water-cooled plate, augmenting the traditional solder joint thermal path. Four Watts per chip or 300 to 400 Watts for a 100-chip module can be dissipated.

Finally, liquid immersion cooling (80) and cryogenic applications (17), (18) have also been demonstrated for C4 interconnected structures. The latter application, for Josephson devices that operate at 4.2 K, was especially noteworthy, because all of the materials were subjected to a very large ΔT between room temperature and operating temperature, and because the intrinsic materials properties, such as resistance to cracking, are quite degraded at low temperatures. Orthogonal solder connections are also used to join silicon slices at right angles to each other. For such a joint, matched expansion materials are required (17), (18).

COMPARISONS OF C4 WITH WIRE BOND AND TAB

ELECTRICAL — The electrical performance of the packaged chip may well be determined by the mode of chip interconnection (81). This arises not from the intrinsic properties of the interconnection but from the package geometry which is, in part, imposed by the chip-level interconnection approach. For example, a wire-bonded silicon device is normally associated with a leadframe, which for fabrication reasons or board-level interconnection geometry requirements, may require rather long leads, well separated from a ground or reference plane. (Alternative wire-bonded structures, such as pin grid arrays or direct-chip attach to a substrate, may provide a much different electrical signal environment.) High-end computer packaging designers are aware of the electrical performance issues associated with packaging, usually for their own customized applications. For this reason, as well as because low-performance systems are not so often package limited, focus on mid-range structures that are frequently gated by package configuration and performance. To be considered are chips having 84 and 180 I/Os packaged using each of the three primary chip interconnection approaches: C4 solder-joining, wire bonding, and TAB (tape-automated-bonding).

The packages to be compared are: (1) a C4-joined, metallized ceramic single-chip carrier having swaged pins; (2) a C4-joined, multilayer ceramic, single-chip carrier having brazed pins; (3) a wire-bonded pin grid array; (4) a wire-bonded plastic leaded chip carrier (PLCC) (84 I/O only); and (5) a TAB-packaged device. The pin grid for (1), (2), and (3) is fixed at 2.54 mm; the PLCC is fixed at 1.27 mm lead pitch; the TAB package is fixed at 0.5 mm lead

pitch. (Note, package (2) is a somewhat artificial configuration, given that usually more than one C4-joined chip is joined to an MLC module.) The objective here is to give a rough basis for comparing package performance, using the lead geometry of the packages. The basis for comparison is the calculated range in the value of lead inductance. This provides a starting point for estimating the effective package inductance, L_{eff}. This quantity is used to predict the inductance-based voltage swing (or ΔI noise), $\Delta V = L_{eff}\, n(dI/dt)$, where n is the number of simultaneously switching drivers, having a given current surge dI/dt Shown in Fig. 20 is the lead inductance calculated for various packages using a three-dimensional inductance program (82). Lead inductance may be used to estimate the effective package inductance, given a detailed knowledge of chip configuration and the number and location of power and ground leads (all of which should be optimized for the package and chip of interest (83)). The relation between lead inductance and L_{eff} is quite geometry-specific and will not be expanded upon further. Signal coupled noise, which is equally dependent on a detailed package description, will also not be considered here. Referring to the table of package inductances, it should be noted that for most of the package geometries chosen, a broad range of lead inductance is observed, indicating the need for personalization of signal and power/ground assignments.

DENSITY — A dramatic difference in attainable connection density between area arrays and peripheral pads was shown previously in (Fig. 11) and Fig. 12. With the advent of area TAB and multiple wire-bond layers, improvements in density are being made (Fig. 21). High performance and I/O technologies in both leading edge CMOS and bipolar chips will favor area array technologies over peripheral. The present area of applicability of wire bond, TAB, and C4 is shown in Fig. 22.

PROCESS — A typical process comparison of C4, wire-bond, and TAB for a multichip module application is shown in Fig. 23. Although process costs and density are major influencing factors, reliability, component and tooling costs, thermal performance, hermeticity requirements, need for reworking, turnaround time, and electrical parameters have all been shown to play major roles in the selection process. While reworking is easy for C4 interconnections, inspectability of the joints is not. Although wire bonding and TAB do not restrict chip sizes and thermal expansion coefficients of the substrates, these restrictions are inconsequential for C4 when matched expansion substrates are used. High thermal capability options have been shown for all of the technologies.

Fig. 20. Calculated Lead Inductance For Single Chip Modules (82)

	C4[1]	C4[2]	WB[1]	WB[4]	TAB[5]
84 I/O Module					
Substrate Size (mm)	24	24	24	30	18
Lead Length (mm)					
(min.)	2	1	2	12.5	6.5
(max.)	17	13.5	17	17.7	9.0
Lead Inductance (nH)					
(min.)	3.5	2.5	7.0	16.0	5.0
(max.)	13.3	6.3	19.0	23.0	7.2
180 I/O Module					
Substrate Size (mm)	36	36	36	—	30
Lead Length (mm)					
(min.)	2	1	2	—	11.3
(max.)	23	20	23	—	16.3
Lead Inductance (nH)					
(min.)	3.5	2.5	7.0	—	8.5
(max.)	17	8.2	24	—	12.5

Note

1. C4 Bonded, MLC Binded Pin Grid Array
2. C4 Bonded, MLC Pin Grid Array
3. Wire-Bonded Pin Grid Array
4. Wire-Bonded Plastic Leaded Chip Carrier
5. Tape-Automated-Bonded Package

The choice of assembly technique also depends upon constraints not discussed in the text so far. A manufacturer building an assembly for in-house use may opt for a technology for which he or she already has manufacturing capability, even though that technology may not be optimized for performance. Likewise, a small company relying completely on vendors for chips and substrates must choose from what is available rather than what is most desirable. This leads to the conclusion that while technical trade-off discussion is desirable, the decision to use one technology over another is more than technical.

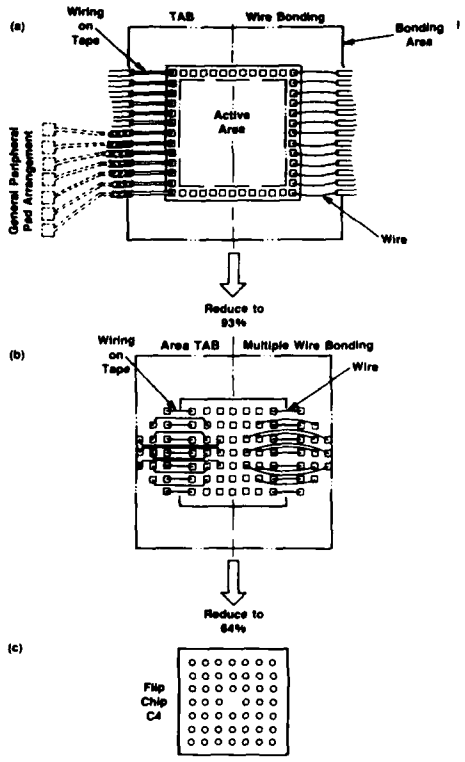


Fig. 21. Relationships Among Various Bonding Methods and Bonding Space
(After K. Ohtsuka, Ref. (84), 1984 Nikkei Electronics Microdevices).

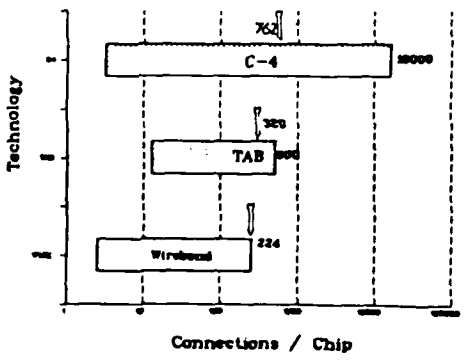


Fig. 22. Range of Applicability
(After R. Tummala and E. Rymaszewski, Ref. (81), 1988.)

Fig. 23. Chip Interconnection Assembly Comparison for Multichip Module.
(After R. Tummala and E. Rymaszewski, Ref. (81), 1988.)

	C4		WIRE BOND		TAB
1.	Flux site	1.	Solder or epoxy preform place	1.	Reel tape into bond position
2.	Align and place chip(s)	2.	Align and place chip	2.	Align and place chip
3.	Reflow to Bond all Pads (On all chips)	3.	Die bond	3.	Inner-lead-bond (One-chip-at-a-time)
4.	Clean flux	4.	Wire-bond each wire on chip and S/S (One-wire-at-a-time)	4.	Encapsulate
5.	Test	5.	Test	5.	Test-burn-in (optional)
6.	Encapsulate Finish module assembly	6.	Encapsulate Finish module assembly	6.	Excise chip
				7.	Align and place assembly on S/S
				8.	Outer-lead-bond
				9.	Test

C4 — FUTURE TRENDS

As VLSI proceeds to denser chips, C4 densification will follow. Bumps of 25 microns, resulting in over 10,000 pads per chip in experimental devices, have already been shown (16). Photolithographic processing will be a must for such devices, and with that area arrays will become more and more common. Significant activity in matched-expansion substrates, or alternatives that alleviate the fatigue limitations, is being actively pursued. High thermal conductivity packaging materials and innovative configurations to improve heat dissipation of the devices are also being evaluated. More attention will be given to defects, for both yield and reliability needs. Efforts to purify all packaging materials for low alpha emission will continue.

The range of applications for C4-like structures is becoming much wider, given the more diverse materials used in new electronic devices and substrates (for example, In C4's on *GaAs* chips on sapphire substrates for optoelectronics).

As the competitive position and capability of solder-bump interconnections become markedly improved with these recent developments, it is likely that C4s will be more generally used in the future.

REFERENCES

1. M. Ohshima, A. Kenmotsu, and I. Ishi, *The Second International Electronics Packaging Conference* (1982).

2. L.J. Fried, J. Havas, J.S. Lechaton, J.S. Logan, G. Paal, and P. Totta, *IBM J. Res. Develop.***26**, 362–371 (1982).
3. L.S. Goldmann and P.A. Totta, *Solid State Tech.* (1983).
4. S. Oktay and H.C. Kammerer, *IBM J. Res. Develop.***26**, 55–66 (1982).
5. P.A. Totta and R.P. Sopher, *IBM J. Res. Develop.***5**, 226–238 (1969).
6. L.F. Miller, *IBM J. Res. Develop.***13**, 239–250 (1969).
7. P.A. Totta, *21st Electronics Components Conf.*, 275 (1971).
8. A.J. Blodgett, *IEEE Tr. Components***CHMT-3**, 634 (1980).
9. B.T. Clark and Y.M. Hill, *IEEE Tr. Components***CHMT-3**, 89 (1980).
10. B.T. Clark, *IEEE Tr. Components***CHMT-4** (1981).
11. A.J. Blodgett and D.R. Barbour, *IBM J. Res. Develop.***26**, 30–36 (1982).
12. D.R. Barbour, S. Oktay, and R.A. Rinne, *Electronics* (1982).
13. J.P. Dietz, *IBM Press Release* (1986).
14. T. Kamei and M. Nakamura, *28th Electronic Comp. Conf. Proc.*, 172–182 (1978).
15. P. Schmid and H. Melchior, *Rev. Sci Inst.***55**, 1854 (1984).
16. W. Weston, *4th Annual Microelctronic Interconnect Conf.* (1985).
17. S.K. Lahiri, H.R. Bickford, P. Geldermans, K.R. Grebe, and P.A. Moskowitz, *IEEE Tr. Comp.***CHMT-5**, 271 (1982).
18. C.Y. Ting, K. Grebe, and D. Waldman, *J. Electrochem. Soc.***129**, 859–864 (1982).
19. K.R. Grebe, *J. Electrochem. Soc.* (1980).
20. S.E. Greer, *28th Electronic Comp. Conf. Proc.*, 166–171 (1978).
21. I. Lang, *Electro. Science and Tech.***9**, 296 (1982).
22. J.D. Leibowitz and W.E. Winters, *IEPC* (1982).
23. J. Fisher, *ECC*, 219 (1984).
24. R.K. Wassick, *Soldering in Electronics*, (Electrochemical Publications Limited, 1984).
25. D.J. Bendz, R.W. Gedney, and J. Rasile, *IBM J. Res. Develop.***26**, 278–285 (1982).
26. J. Gow 3rd and A.R. Leach, *Circuits Mfg.*, 72 (1983).
27. T. Takenaka, F. Kobayashi, and T. Netsu, *1984 ISHM Conf. Proc.*, 419 (1984).
28. F. Kobayashi, T. Takenaka, M. Ishida, and T. Netsu, *Proc. 1982 IMC*, 425 (1982).
29. K.C. Norris and A.H. Landzberg, *IBM J. Res. Develop.***13**, 266–271 (1969).
30. V.K. Nagesh, *Rel. Phys. Symp.* (1982).
31. D.L. Brownell and G.C. Waite, *1974 ISHM Conf. Proc.*, 77 (1974).
32. W. Roush, *Proc. 2nd Int'l Conf. on Behavior of Materials*, 780 (1976).
33. G. Clatterbaugh and H. Charles, *ISHM*, 31 (1985).
34. T. Yamada, K. Otsutani, K. Otsukani, K. Sahara, and K. Otsuka, *IEPS*, 551–557 (1985).
35. H. Nye, Private Communication, 1986
36. H. Curtis, *IBM System 38 Technical Development* (1978).
37. L.S. Goldmann, *Proceedings 22nd Electronic Comp. Conf.*, 332 (1972).
38. T. Kawanobe, K. Miyamoto, and Y. Inaba, *31st Electronic Comp. Conf. Proc.*, 149 (1981).
39. T.R. Myers, *Proc. ECC*, 131 (1969).
40. J. Marshall and F. Rode, *Solid State Tech.*, 87 (1979).
41. N. Nuzaki, K. Nakamura, and I. Taubokawa, *28th Electronic Comp. Conf. Proc.*, 183 (1977).
42. E. Rasmanis, *1973 ISHM Conf. Proc.*, 3B-3–1 (1973).
43. E. Anderson, *Microelectronic Interconnect Conference* (1984).
44. A. VanDer Drift, W.G. Gelling, and A. Rademakers, *Phillips Technical Review***34**, 85 (1974).
45. N. Tsujimoto, *Electro. Parts***19**, 51 (1980).
46. R.T. Howard, *IBM J. Res. Develop.***26**, 372–389 (1982).
47. K.J. Puttlitz, *Journal of Electronic Materials***13**, 29–46 (1984).
48. S. Teed and V. Marcotte, *NATAS*, 77 (1981).
49. J. Temmyo, K. Aoki, H. Yoshikiyo, S. Tsurumi, and Y. Takeuchi, *J. Appl. Phys.***54**, 5282 (1983).
50. D.A. Jeannotte, *Electronic Comp. Conf. Proc.*, 334 (1969).
51. L.F. Coffin, *Appl. Mechanics Res.***1**, 129–141 (1962).
52. P.A. Tobias, N.A. Sinclair, and A.S. Van, *ISHM Proc.*, 360 (1976).
53. H.J. Shah and J.H. Kelly, *1970 ISHM Proc.* (1970), (Paper 3.4).
54. N.A. Sinclair, *Int'l Electrical Proc.*, 56 (1982).
55. E. Levine and J. Ordonez, *31st Electronic Comp. Conf. Proc.*, 515–519 (1981).
56. L.S. Goldmann, *IBM J. Res. Develop***3**, 251–265 (1969).
57. R.T. Howard, S.W. Sobeck, and C. Sanetra, *Solid State Tech.***26**, 115–122 (1983).
58. R. Satoh, M. Ohshima, H. Komura, I. Ishi, and K. Serizawa, *IEPS*, 455 (1983).
59. N. Matsui, S. Sasaki, and T. Ohsaki, *Proc. IEEE 37th Electronic Components Conf.*, 573–578 (1987).
60. E. Horikoshi, K. Hoshimoto, T. Sato, and H. Nakajima, *Japanese Patent No. 62-117346* (1987).
61. P.M. Hall, *31st Electronic Comp. Conf. Proc.*, 172–180 (1981).
62. V.D. Coombs, *26th Electronic Comp. Conf. Proc.*, 67–72 (1976).
63. G. Cherian, *IEPS*, 710 (1984).

64. T. Watari and H. Murano, *Proceedings of the 35th ECC* (1985).
65. L.S. Goldmann, R.J. Herdizk, N.G. Koopman, and V.C. Marcotte, *ECC Proceedings***27**, 25 (1977).
66. R.T. Howard, *32nd Electronic Comp. Conf. Proc.*, 367–384 (1982).
67. W. Roush and J. Jaspal, *32nd Electronic Comp. Conf. Proc.*, 342 (1982).
68. J.H. Lau and D.W. Rice, *Solid State Tech.*, 91 (1985).
69. W. Engelmaier, *IEPS*, 15 (1982).
70. V. Marcotte, M. Ricker, and N. Koopman, *IMS/ASM/AWS Symposium* (1985).
71. V.C. Marcotte and N.G. Koopman, *ECC*, 157–164 (1981).
72. D. Bouldin, *J. Of Electronic Materials***10**, 747 (1981).
73. R. Howard, *ECC*, 453 (1984).
74. S. Oktay, *IBM J. Res. Develop.***13**, 272–285 (1969).
75. R.C. Buchanan and M.D. Reeber, *Solid State Technology***26**, 39 (1973).
76. D.M. Cavanaugh, *ISHM Proc.*, 214 (1975).
77. V. Kurokawa, K. Utsumi, H. Takamizawa, T. Kamata, and S. Noguchi, *IEEE TR.***CHMT-8**, 247 (1985).
78. K. Okutani, K. Otsuka, K. Sahara, and K. Satoh, *IEPS*, 299 (1984).
79. R.C. Chu, U.P. Hwang, and R.E. Simons, *IBM J. Res. Develop.***26**, 45–54 (1982).
80. R.E. Simmons and R.C. Chu, *ISHM*, 314 (1985).
81. R.R. Tummala and E.J. Rymaszewski, *Microelectronics Packaging Handbook*, (Van Nostrand Reinhold, New York, 1988), (To be published).
82. M.F. Bregman, Private Communication
83. G.A. Katopis, *Proc. IEEE*, 1405–1415 (1985).
84. K. Ohtsuka, *Nikkei Electronics Microdevices 2* (1984).

SURFACE MOUNT TECHNOLOGY

Jim Walker
National Semiconductor
Santa Clara, California, USA

ABSTRACT

Surface mount technology provides many benefits. Increased production efficiency, smaller size, improved reliability, lower weight, reduced costs are some--just to name a few. Despite these benefits, the transition to surface mount technology in the United States has proceeded at a pace slower than predicted by most industry analysts. The situation has now changed. Manufacturers are switching to surface mount. Worldwide competition dictates they have little or no choice but to implement the technology. This will require changes in assembly, equipment, technology, and circuit board design. Each of these topics will be discussed.

SURFACE MOUNT PACKAGES

Surface mount technology, or SMT, had its beginnings in the late 1960's as part of the U.S. space effort to place a man on the moon. This effort created the hybrid industry, and led to the development of surface mount ceramic packages with low profile and high reliability. The ceramic "Flat Pack" was probably the first surface mount I.C. package developed. Soon thereafter, leadless ceramic chip carriers (LCC) were designed for increased density and smaller board footprint. Both of these packages, being ceramic, were high in cost and initially used in low volume applications.

In the early 1970's, the electronic watch and calculator industry utilized hybrid chip and wire techniques to pack several functions in the small area of a watch case. As volume increased, problems of manual assembly and rework increased costs. A search for solutions led to the creation of alternate component packages for high volume applications.

One such package, developed by Phillips, was the Small Outline Package (S.O.). Intended for the Swiss watch industry (S.O. originally meant Swiss Outline), it soon was adapted to many consumer products built by the Japanese. This has resulted in the S.O. package becoming widely used today, containing most of the lower pin count logic and analog functions prevalent in consumer applications.

The Plastic Leaded Chip Carrier (designated as PCC or PLCC) was developed as a lower cost version of the Ceramic Chip Carrier. The initial purpose of this package was to fit the same sockets and board footprint as the ceramic leadless version. However, as surface mount technology has intensified, the PLCC has now replaced many applications where the LCC had been used. Since the PLCC has leads on all four sides, it has become the major package of choice for pin-counts in the range of 28 to 84 leads.

As silicon technology pushes for more and more functions on a chip, this also necessitates a corresponding change in packaging technology. The S.O. and PLCC are effective for leads up to 84. Above this number, the packages become too large (being on 50 mil centers) to maintain the density needed. In addition, as packages become larger, they become more difficult to solder to the circuit board and meet reliability standards. So, new surface mount packages must be developed.

The Plastic Quad Flatpack (PQFP) has been developed to meet the increased demand for reliable packages above 84 leads. There are several types of Quad Flatpacks ranging up to 244 leads and made by both the U.S. and Japanese.

Table 1 lists the most common surface mount packages used today for integrated circuits.

TECHNOLOGY CHANGES FOR SURFACE MOUNT

One of the most significant changes when switching to surface mount is soldering technology. Existing through-hole processes with the dual-in-line (DIP) package use wave soldering as the mainstream board attachment method. Boards are moved over a flowing wave of hot, molten solder with the bottom side the recipient of transferred solder. Using this method with surface mount components can cause many problems specific to I.C. packages. Shadowing can occur behind chip carriers and other large components, or when one component is placed too close to another. Long term reliability can be compromised when the Small Outline is subjected to excessive wave solder immersion time.

Reflow soldering is the preferred method for surface mounting. The reflow process involves the use of a solder paste (tiny balls of solder suspended in an organic medium--similar in consistancy to thick paint). The paste is applied by silk or stencil screening to metal pads (sometimes referred to as "lands") on the circuit board. The device is then placed on the solder-coated pads. The board is heated until the paste "reflows" or melts around the leads of the device, and is subsequently cooled. Since parts are mounted on the surface of the circuit board, the board may be inverted and the same procedure used on the opposite side. In this manner, components are mounted on both sides, thus doubling the density over single sided through-hole assembly. A comparison of through-hole versus surface mount process flow is shown in Figure 1.

SOLDER PASTE APPLICATION

As mentioned above, solder paste must be suitable for application by stencil or silkscreen. It should be resistant to solder balling, and reflow at the proper temperature. The flux residue must be easily removable. Tin-lead or tin-lead-silver solder paste compositions are the most common types. A solids content of 88-90 % is preferred.

The shape of the solder particles is an important factor. They should be spherical in shape with uniform diameters and a minimum amount of elongation. The size distribution should be uniform, as uneven distribution causes improper melting and subsequent expulsion of smaller solder balls away from their proper sites.

The paste can be applied to the board using a stainless steel, wire mesh screen stencilled with an emulsion image of the substrate pads. It is forced through the screen by a V-shaped plastic squeegee in a sweeping manner to the board placed beneath the screen.

ASSEMBLY EQUIPMENT

To optimize the surface mount process, new equipment is required. The equipment used for surface mounting in comparison to that of the through-hole method is shown in Figure 2. Surface mounting utilizes screening, pick-and-place, and solder reflow equipment, all which are not part of the through-hole (insertion) process.

There are various methods for reflowing the solder paste. Vapor phase is currently the most popular and consistent method. It offers precise temperature control through chemical thermodynamics which guarantees that the soldered assembly never reaches a temperature above that of the boiling fluid. The liquid must have a boilnig point minimum of 20 deg C above the reflow temperature of the solder, be non-flammable, non-toxic, have high vapor density and must not decompose or leave a residue. Liquids which meet these criteria generally have boiling temperatures in the 210-230 deg C range, are based upon carbon-fluorine chemistry and are very expensive.

The vapor phase process involves immersing the printed circuit board assembly in the vapor of the fluorinert liquid described above. The vapor condenses on the board and transfers its heat to the board and causes the solder paste to reflow.

Another method of soldering uses absorption of radiant energy to heat up the materials. In infrared reflow soldering, the air in the heating path is not directly heated. There is no heat transfer by direct physical contact. In the process, the board passes into a preheat zone where radiant energy heats the solder paste to volatilize solvents and avoid thermal shcoking the parts and the board. After preheating, the assembly enters the main heating area where it is raised to reflow temperature and then cooled.

Laser reflow soldering is primarily used for lower volume, high density boards or heat sensitive components. The heat is localized on each solder joint individually by a focused laser beam, and does not contact the main body of the component. This method can apply heat to an area with a diameter of less than 5 mils and typically solders only about 5 joints per second.

PRINTED CIRCUIT BOARD CHANGES

Not only does the process, equipment and technology change when converting to surface mount, the printed circuit board must also change. More area is available for trace routing and component placement due to the absence of the large diameter component lead mounting holes which were required for the through-hole process. Required now are only small diameter via holes (0.013-0.020 inch) between circuit layers.

The design of the board becomes an important factor in that it should be process oriented. It must reflect the handling and material test methods. As an example, the board must be designed to minimize board warpage during the soldering process, as this might result in open solder joints.

FACTORY AUTOMATION

As the changes to convert to surface mount are implemented, one benefit becomes apparent: factory automation. Pick-and-place equipment capabilities are much faster than the DIP inserters. Smaller packages that sit on the board surface don't require complex finger mechanisms to push and guide leads into holes.

Tape and reel delivery systems for IC packages have been developed and standardized. They offer a means to postitively move and index the components at the placement head dependent upon machine speed. Tape and reel allows much greater volumes to be loaded. A 13 inch reel of S.O.-14 lead components contains 2500 pieces whereas 50 are contained in a standard SO tube (rail). This results in a machine loaded with taped components yielding a possible labor savings of 50 to 1 over the tube feeding method.

SMD TODAY

Several industry marketplaces are using surface mount today and will continue if they are to remain competitive. Those markets include:

Consumer Electronics- VCR's, camcorders, cameras, tape players.

Computers- Lap top, disk drives, memory modules and arrays

Telecommunications- mobile phones, microwave devices

Military- Satellite, rockets, avionics

SUMMARY

The initial drawbacks to surface mount have been overcome. Design and manufacturing expertise has been developed. Component price is equivalent to the DIP in large volumes. Assembly equipment is now available from many suppliers. Virtually all electronic market segments are examining surface mount technology.

TABLE 1

S.M.D. PACKAGES FOR INTEGRATED CIRCUITS

(EIA) J E D E C

PINS	S.O NARROW	S.O WIDE	P L C C SQUARE	P L C C RECTANGLE	P Q F P	TAPEPAK
8	X					
14	X	X				
16	X	X				
18		X		X		
20		X	X			
24		X				
28		X	X			
32				X		
40						X
44			X			
52			X		X	
68			X			
84			X		X	X
100			X		X	
124			X			
132					X	X
164					X	
196					X	
244					X	

FIGURE 1. Assembly Process

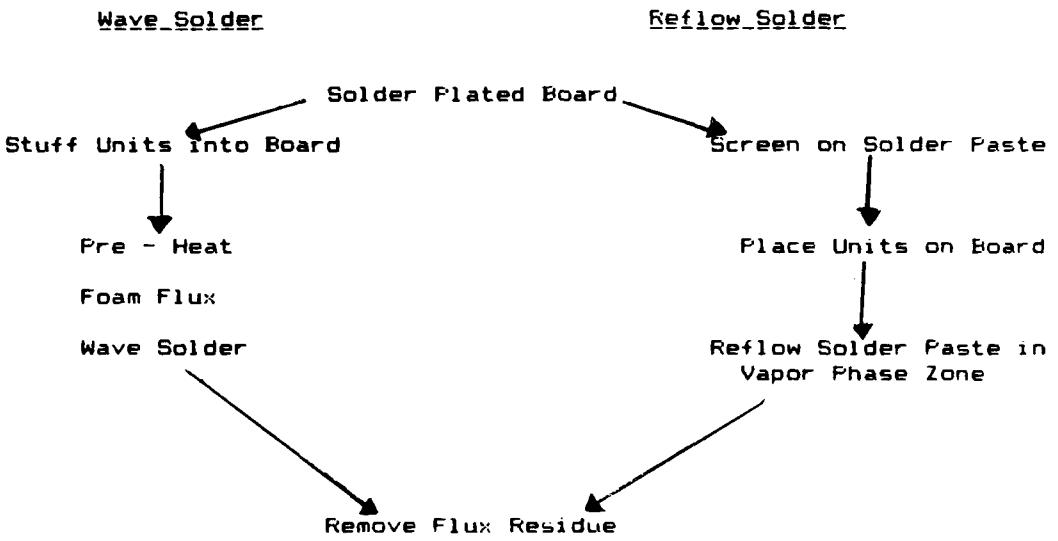


FIGURE 2. Equipment For Assembly

WAVE SOLDER	REFLOW SOLDER
Dip Inserter	Screen Printer
Wave Solder	Placement M/C
Cleaning Station	Vapor Phase Reflow System
	Cleaning Station

MATERIAL CONSIDERATIONS IN CONNECTOR DESIGN

Robert S. Mroczkowski
AMP Incorporated
Harrisburg, Pennsylvania, USA

ABSTRACT

An electronic connector is a system which, in most cases, consists of metallic contact finishes, metallic spring members and polymer housings. The selection of appropriate materials for these functions depends on the design of the connector and application requirements, both operational and environmental. In this paper, material characteristics of particular importance to connector design will be emphasized with respect to their effects on connector performance. These fundamental considerations provide the background against which material/design tradeoffs must be balanced.

AN ELECTRICAL CONNECTOR can be described in terms of its function or its structure. From a functional viewpoint an electrical connector is intended to transmit electrical power, or signals, between two subunits of an electrical or electronic system without introducing unacceptable losses in power or distortion of the signal. This functional requirement can be translated into a requirement that the contact interface between the two halves of the connector has a low resistance and maintains that low resistance during the lifetime of the connector.

In addition to these resistance requirements, there are also durability requirements on the contact interface. In most cases a connector is used so that the subunits can be separated from one another for a variety of reasons. Such separable connectors are intended to provide acceptable performance after some number of mating cycles.

While a long list of connector requirements could be drawn up, the two requirements mentioned, low and stable contact resistance and acceptable durability, are sufficient for our purposes. This paper will provide a brief description of the contact interface and its characteristics to develop an understanding of the relationship between contact interface requirements and the structure of an electrical connector.

Structurally, an electrical connector has been described, not entirely facetiously, as "two contact platings held together by supporting structures". Implicit in this description is the contact interface which exists between the contact platings. This leaves us with four structural components of an electrical connector which are of importance. They are:

1. The contact interface
2. The contact finish
3. The contact spring
4. The connector housing.

Fig. 1 - Cross section of a separable connector. (18X)

Figure 1 contains a photomicrograph of a longitudinal section of a separable connector. The left half of the connector pair is a post which is typically soldered into a printed circuit board. The right half is a receptacle which is terminated to a wire, in this case by a crimp. The contact interface is formed at the point where the contact beams, the contact

springs, of the receptacle mate to the post. In this example, the contact finish for both the post and the contact beams is gold over nickel. The base metal of the contact springs is phosphor bronze. The connector housing is also visible in Figure 1.

Each of these functions has its unique characteristics and requirements. In this paper an overview of the materials requirements and considerations which go into the functional and structural design of an electrical connector will be presented. The discussion is limited in scope. Not all materials or design options are explored, and the discussion is qualitative rather than quantitative. The intent is to provide a basic framework to support the claim that a connector is a system in which the materials/design interactions must be considered in order to provide the desired performance.

REQUIREMENTS FOR A CONTACT INTERFACE

Two requirements on the contact interface have been mentioned, low resistance and durability. In this section only the resistance requirement will be discussed. The discussion on durability will be deferred to the section on contact finishes. To understand the source of contact resistance the structure, and microstructure, of the contact interface must be considered.

A graphic analogy (2) suggests that a typical contact interface, if magnified about 300 million times, would look much like Vermont on top of New Hampshire! A much less dramatic description is provided in Figure 2.

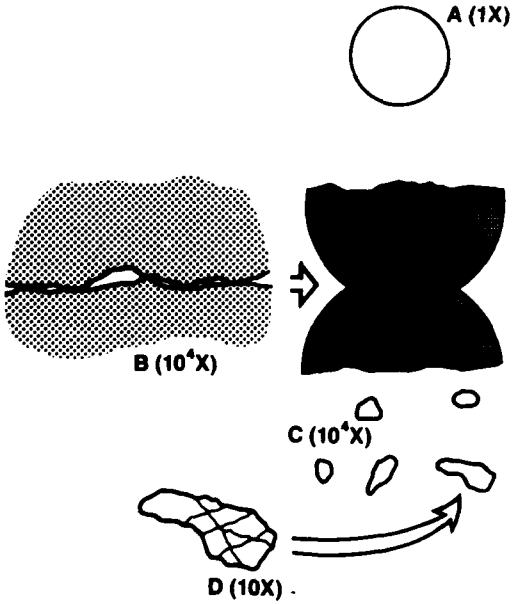


Fig. 2 - Schematic representation of a contact interface.

Consider two spheres in contact under a load, L. The macroscopic contact area, for this geometry, will be circular as shown in Figure 2a. In actuality, due to the inherent roughness of all surfaces at the microscopic level at which contact occurs, only high spots on the surface will be in contact as shown in Figure 2b. This arrangement leads to the view in Figure 2c. At this point it is apparent that the true contact area is only a small fraction of the macroscopic contact area - less than a percent to a few percent depending on the applied load. But we are not done yet. The area discussed to this point is the area in mechanical contact. In cases where the surfaces contain films, usually oxides, the mechanical contact may be electrically insulated by the oxides. Electrical contact occurs only in areas of metallic contact. It can be shown (3) that when oxide covered surfaces deform against one another the oxides crack and separate. Metallic contact occurs only where cracks in the two oxide covered surfaces are coincident. Figure 2d schematically illustrates areas of electrical contact as cracks in the surface film.

Experimental verification (2) of the validity of this representation is shown in Figure 3.

Fig. 3 - Back Scatter SEM micrograph of a contact spot on aluminum (4000X).

In this case the contact was to a flat plate of aluminum with its intrinsic surface oxide. Note that the surface oxides have cracked and separated during deformation. Note the "walls" of aluminum which have extruded through the cracks in the oxide. It is these "walls" which would be the areas of metallic, therefore, electrical contact.

This description needs further elaboration to accurately represent what occurs in the mating of typical separable connectors. One important factor which has not been mentioned is the sliding action which occurs on mating, which is referred to as contact wipe. Wipe is effective in displacing surface

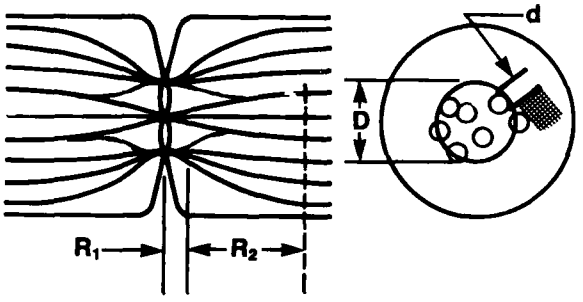
contaminants in addition to disrupting oxide films (4.5). In this manner the area of electrical contact can be significantly increased. The tangential stresses at the contact interface as wiping begins also bring about an increase in the contact area (6). It remains, however, a small fraction of the macroscopic contact area.

A second factor of importance is the effect of the distribution of contact spots on the contact resistance of the interface. For a detailed discussion of this effect see reference 7. For our purposes a simplified discussion, assuming film free surfaces, will suffice.

At each individual spot there exists a resistance, called constriction resistance, which is due to the fact that current flow is constricted to flow through the individual points of contact rather than being able to use the full conductor cross section. Constriction resistance is a purely geometric effect, and does not depend on the physical properties of the contact interface. The magnitude of the constriction resistance at any contact spot is given by:

$$\rho/d \qquad (1)$$

where ρ is the resistivity of the material and d is the diameter of the contact spot (8,9). In a typical contact interface there will be a number of such contact spots, schematically illustrated in Figure 4, all of which will be electricaly in parallel.



$$R_{CONTACT} = R_1 + R_2 = \rho/nd + \rho/D$$

Fig. 4 - Constriction resistance for multiple contact regions. Individual contact spots have an effective diameter, d, and the overall spot distribution a diameter, D.

The resulting overall constriction resistance is given by the series combination of these individual contact spot constriction resistances and the constriction resistance due to the distribution of the contact spots at the interface. In this case, the following equation applies;

$$Rc = \rho/nd + \rho/D \qquad (2)$$

where n is the number of spots, d is the average diameter of the individual spots, and D is the diameter of a circle containing the distribution of contact spots (9).

Constriction resistance will decrease as the normal force which produces the contact interface increases. The decrease will result from the fact that n and d will both increase as the normal force increases. After a sufficient number of contact spots have been created, the constriction resistance will be dominated by the distributional term ρ/D. Under these conditions the contact will function as though the full distributional area was conducting. D will also increase, slowly, with increasing normal force.

Constriction resistance versus normal force, then, will have a relationship of the form shown in Figure 5.

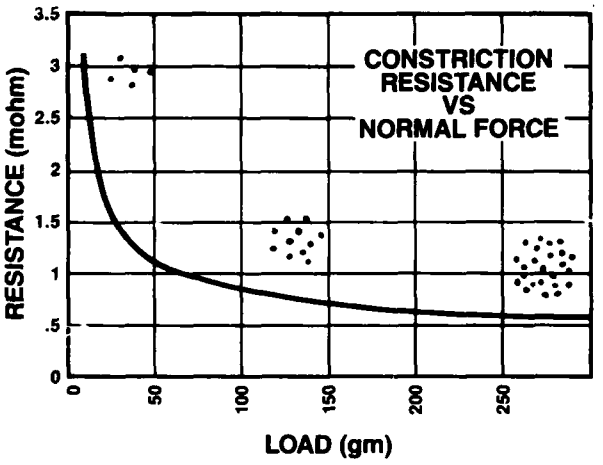


Fig. 5 - Schematic representation of contact resistance versus normal force illustrating the effect of contact spot distribution.

Constriction resistance decreases rapidly initially due to the increase in the number of contact spots. When the number of contact spots is in the range of ten to twenty, the distribution term becomes dominant and resistance decreases more slowly with normal force. The rate of decrease becomes proportional to the hardness of the contact material since it is the hardness which determines the distributional spot size at a given applied load. In fact, under these so called "full area" contact conditions the constriction resistance can be approximated by (10):

$$Rc \sim \rho \sqrt{H/F} \qquad (3)$$

where H is the hardness and F the contact normal force.

Equation 3, and Figure 5, demonstrate one reason contact normal force is an important connector design parameter. The normal force

applied establishes the minimum contact resistance which can be realized at the contact interface for a given material.

As mentioned previously, up to this point the surfaces in contact have been assumed to be clean, that is, free from surface films or contaminants. This is not always the case, as discussed with reference to Figure 2. Another important function of normal force, in combination with wiping action, is in disrupting films (Figure 3) and displacing surface contaminants (4,5).

A third function of normal force is to provide mechanical stability to the contact system. Such stability is achieved by the friction forces which accompany the normal force. These frictional forces reduce the likelihood of motion of the contact interface and, therefore, disturbance of the contact spots with a possible resultant increase in contact resistance. Due to the high redundancy provided by the numerous contact spots, disturbances may not, at least initially, result in resistance increases. Motion of the contact interface is the major degradation mechanism for tin contacts, as will be discussed later. But in any event, disruptive motions of the contact interface should be minimized and increasing normal force is one way to accomplish this end.

This discussion has shown that normal force is an intrinsic parameter of connector design. The value of normal force which is required depends on the contact geometry, the contact finish and the design of the connector; in other words, on the connector system. In this paper we will consider only the effects of contact finish. For a discussion of other normal force considerations see references 11 and 12.

REQUIREMENTS FOR A CONTACT FINISH

A contact finish is intended to perform two major functions:

1) to provide corrosion protection for the base metal contact spring

2) to optimize contact surface characteristics.

The first, corrosion protection, is straightforward. The copper alloys typically used as contact springs are subject to corrosion in many operating environments. The contact finish, in effect, seals off the spring from the environment. This is an important, but not obvious function of a contact finish.

To provide corrosion protection the contact finish must be able to survive application conditions such as mating operations. The durability characteristics and requirements of a contact finish, therefore, become a concern. Durability will depend on the following factors:

1. Contact finish
2. Normal force
3. Contact geometry
4. Mating distance
5. Mating cycles

For this discussion we will consider only the contact finish effects. Two finish characteristics which influence durability are the hardness and coefficient of friction. As hardness increases and coefficient of friction decreases the durability of the finish will improve under any combination of the other factors listed. Approximate hardness and coefficient of friction values are provided for comparison purposes in Table I. The complexity of interactions of the listed parameters in determining the durability of a contact places discussion of this issue out of the scope of this paper. For a review of contact finish wear see reference 13.

Optimization of contact surface characteristics can be realized through two different effects depending on whether the contact finish is a noble or non-noble metal. First, the contact surface can be made a noble metal in which case the probability of surface corrosion films is reduced. Second, the contact surface can be made one in which surface films are readily disrupted on mating of the contacts. Gold is an example of the first case, and tin of the second. Because of this difference, performance characteristics and design requirements are different for noble and non-noble contact finishes. Each will be discussed separately.

NOBLE METAL FINISHES - Noble, or precious metal, contact finishes include gold and palladium and alloys based on these metals. Their value as contact finishes, as mentioned, is due to their corrosion resistance. Apart from contamination, contact finishes of these materials will show minimal surface film effects in most operating environments. Normal force requirements will be minimized when such finishes are specified. Contact resistance versus normal force data for a gold contact interface is shown in Figure 6 (11).

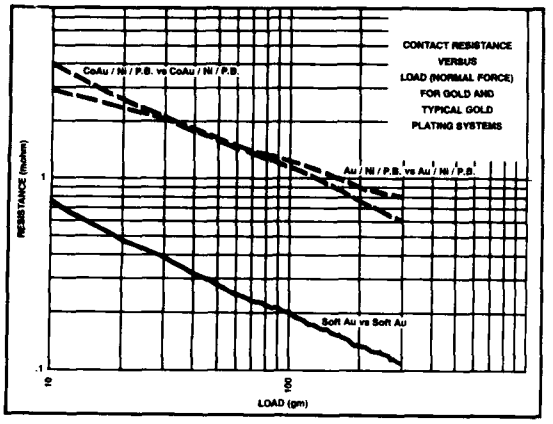


Fig. 6 - Constriction resistance as a function of normal force for electroplated pure gold and cobalt hardened gold (over nickel) contacts on a phosphor bronze substrate.

These data show that a normal force of the order of 25 grams will produce a contact resistance of less than 2 milliohms, a value which is acceptable for signal and low current contacts. Normal forces of this magnitude, however, may not be sufficient to provide the mechanical stability required by a connector without special attention to contact design, as discussed in reference 11.

Precious metal contact finishes usually consist of a combination of layers which are intended to perform different functions. In most cases a nickel underplate will be present for reasons which will be discussed (Also see references 1 and 14). In other words, the contact finish can be viewed as a material system which must be designed in combination with the mechanical system of the connector structure.

The presence of a precious metal plating on a contact surface does not in itself guarantee a film free surface. The plating must be continuous and thick enough to prevent diffusion of base metal constituents of the contact spring from reaching the contact surface. Discontinuities in the plating can result in corrosion at sites where base metal is exposed, and base metal diffusion to the surface can result in surface films. Plating discontinuities can result from different sources depending on the finishing method.

A fundamental source of discontinuities in electroplated contact finishes is porosity, small holes in the plate, which can be explained by examining the kinetics of the electroplating process. All electroplates nucleate at isolated spots on the surface and grow both laterally and vertically producing islands of plating on the surface. Lateral growth eventually produces a continuous layer as the individual islands coalesce. A typical plot of porosity versus plating thickness, for a gold plating, is presented in Figure 7 (15).

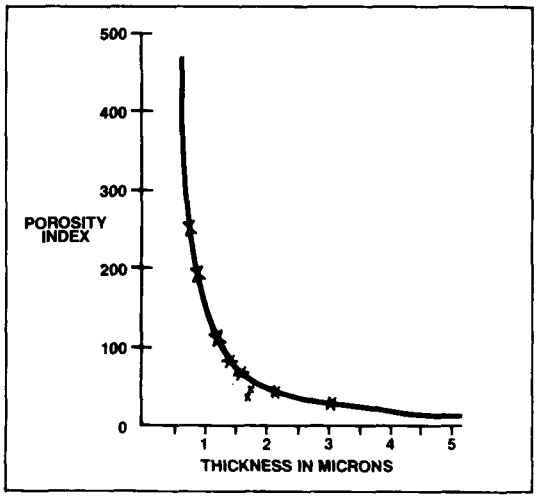


Fig. 7 - Schematic representation of porosity versus plating thickness.

The shape and thickness dependence of such curves depend, of course, on a number of plating parameter and practices. This figure demonstrates one reason why electroplated contact finishes generally lie in the range from 0.6 to 2 microns. Porosity increases rapidly below about 0.6 microns and above 2 microns the porosity is low and the rate of decrease is negligible from a practical viewpoint.

Porosity and other finish defects are, of course, affected by the characteristics of the substrate. Base metal defects, such as inclusions and laminations, and manufacturing defects, such as stamping marks and improper cleaning/handling practices, can also impact the probability of introducing finish defects.

Finish defects are matters of concern due to the fact that they are sites at which corrosion may occur. For example, Figure 8 schematically depicts pore corrosion, Figure 8a, as well as the beneficial effects of a nickel underplate, Figure 8b, in reducing the effects of pore corrosion.

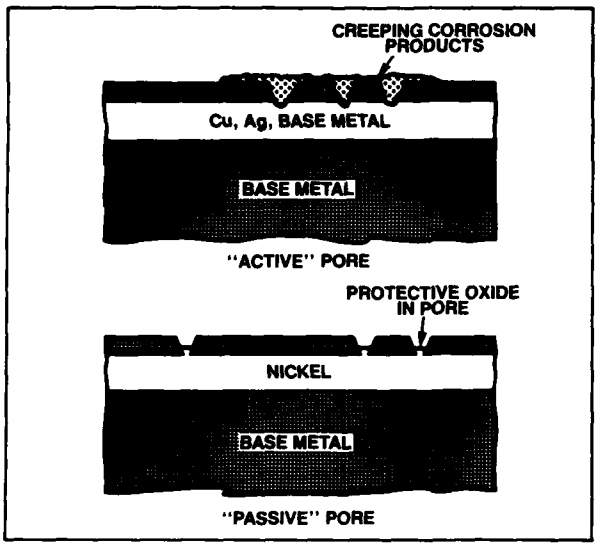


Fig. 8 - Schematic representation of the effects of underplates on pore corrosion.

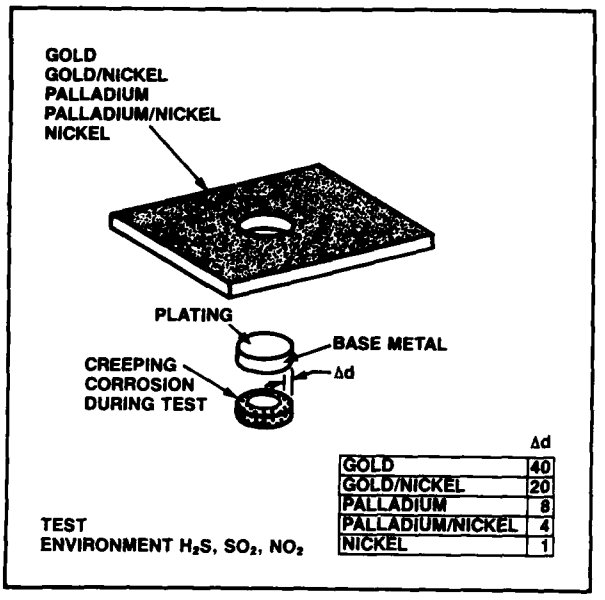


Fig. 9 - Schematic representation of the migration of corrosion products over various contact plating systems.

In Figure 8a, the pore site goes through to a base metal which is subject to corrosion in the operating environment. Corrosion products can volumetrically fill the pore site due to their higher specific volume, or they can migrate up the walls of the pore and reach the surface in that manner. In either case, the presence of corrosion product on the contact surface may lead to degradation of contact resistance. The presence of a nickel underplate, Figure 8b, reduces the effects of pore corrosion since, in typical electronic environments, the exposed nickel will form a limited thickness oxide film which does not fill the pore site, and does not migrate. In effect, the nickel seals off the base of the pore site with respect to detrimental corrosion products. Similar effects will occur at other defect sites.

Another corrosion related benefit of nickel underplates is in reducing corrosion product migration to the contact interface from corrosion sources elsewhere on the contact. For example, contact springs may be plated prior to stamping in which case the stamped contact will have exposed base metal edges which may be in proximity to the contact area. An example of some data demonstrating this effect is presented in Figure 9 (16).

The source of the corrosion is the exposed copper alloy edge resulting from the stamping of the disc from the plated coupon. The test environment, which consists of H_2S, SO_2 and NO_2, is highly corrosive to copper. The copper corrosion products which are created in the test environment can migrate over the surface. From the data of Figure 9 we see that presence of a nickel underplate cuts the migration distance in half for both gold and palladium contact finishes. Note also that the migration rate over nickel is very low. Our interpretation of this data is that the nickel barrier serves to retard the motion of corrosion products from the base metal to the contact surface. But once corrosion products reach the precious metal, migration occurs at a significantly faster rate. Migration over gold occurs more rapidly than over palladium. This effect is believed to be related to the fact that gold surfaces are more noble than the catalytic surface of palladium which tends to promote thin organic surface films. An example of a contact exposed to a similar environment, in this case containing chlorine in addition to the previously listed corrodants, is presented in Figure 10 (17). Both contacts were plated with an overall nickel and a selective gold in the contact area. They differ in that the lower contact also has a gold flash, about a tenth of a micron, over the entire contact.

Fig. 10 - Migration of corrosion products on an edge card contact after exposure to a test environment simulating an industrial exposure.

Migration of the corrosion products which form in the area of the stamped lance, where plating coverage is affected by the roughness of the stamped surface, is minimal over the nickel, but quite extensive over the gold flash. These examples demonstrate that nickel serves as a barrier to corrosion product migration in such environments, which represent a majority of connector operating environments.

A third benefit of nickel underplates is in reducing the effects of base metal diffusion. Figure 11 shows relative diffusion rates of copper through gold, palladium, silver and nickel electroplates (18).

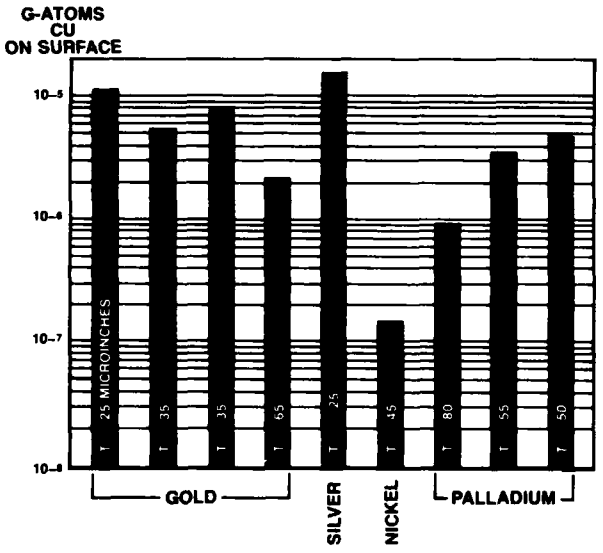


Fig. 11 - Effects of plating materials on the diffusion of copper.

Nickel can be seen to be an effective diffusion barrier to copper. Similar performance occurs with respect to other typical base metal constituents such as zinc and beryllium. In this manner, nickel serves to significantly reduce the probability of base metal constituents getting to the contact surface where they can oxidize.

Last, but not least, nickel underplates improve the durability of gold platings by providing a hard supporting layer beneath the gold. This effect is quite striking at gold thicknesses in the typical contact finish range of 0.25 to 2 microns. Figure 12, adapted from reference 19, shows this improvement dramatically.

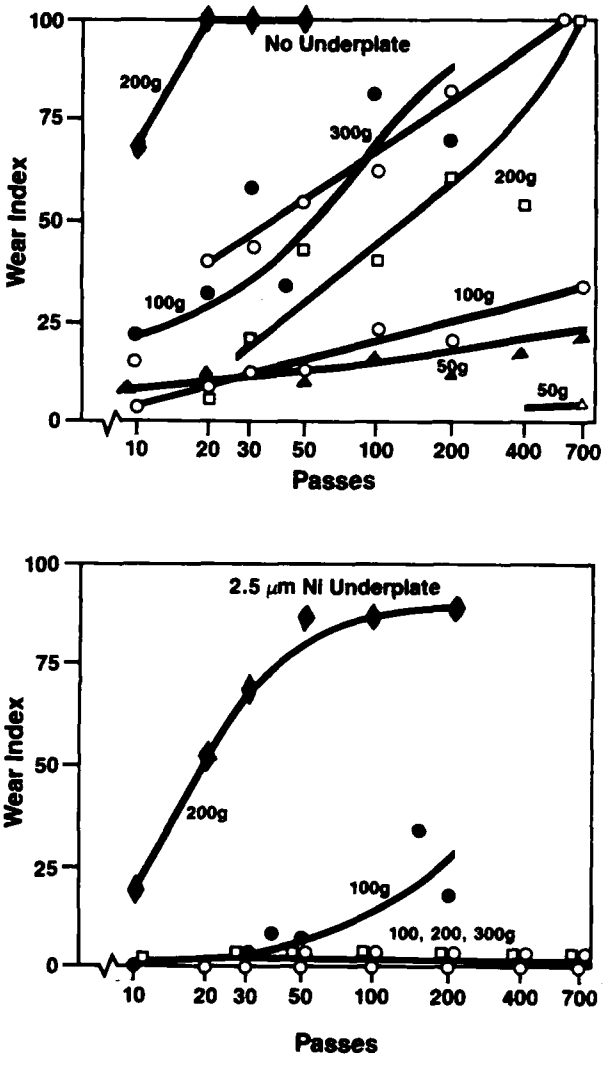


Fig. 12. - Wear data on gold contact surfaces with and without a nickel underplate over copper (solid markers) and beryllium copper (open markers) substrates.

Even with a hard base metal such as beryllium copper, the effect of the nickel underplate is marked.

In summary, precious metal contact platings are a system which take advantage of the following beneficial characteristics of nickel underplates in most connector application environments:

1. Pore corrosion is reduced.
2. Diffusion of base metal constituents is retarded.
3. The durability of precious metal platings is improved.
4. The migration of corrosion products is inhibited.

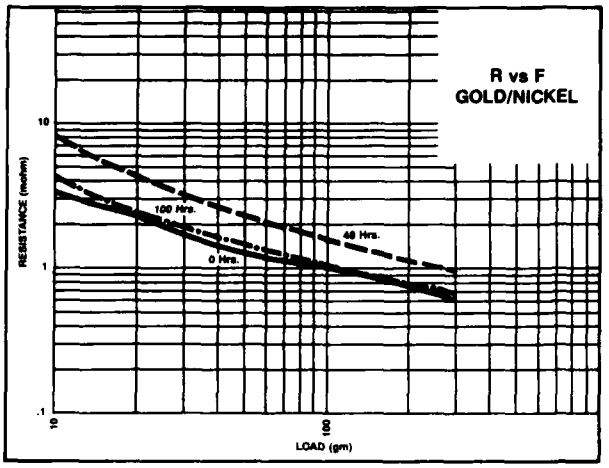


Fig. 13 - Contact resistance versus normal force for a cobalt hardened gold (over nickel) contact surfaces subjected to a test environment simulating an industrial exposure.

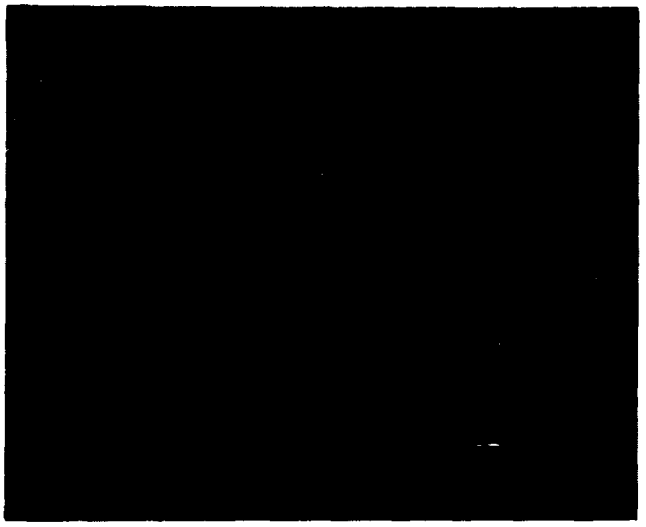


Fig. 14 - Corrosion "haloes" around pore sites on gold over nickel contact surface exposed to a corrosive environment.

Gold and palladium, and alloys based on these metals, are the most common precious metal contact finishes. While both are noble, they do differ in their corrosion resistance. Gold is corrosion resistant in virtually any environment, while palladium has catalytic characteristics and is reactive in strongly oxidizing environments. These differences are reflected in contact performance after environmental exposure as illustrated in Figures 13, 15 and 16 (20). These figures contain contact resistance data from test flats subjected to an environmental exposure which was intended to simulate a 10 year exposure to an "industrial" environment (20). Figure 13 shows the data for gold/nickel platings. There is a slight increase in resistance which reflects the migration of corrosion products from pore sites in the plating. Figure 14 shows spreading of corrosion products around a pore site. Probe points making contact on or near such corrosion products will show increased contact resistance as discussed in reference 14.

The data for palladium, Figure 15, show larger increases indicative of the lower corrosion resistance of palladium compared to gold.

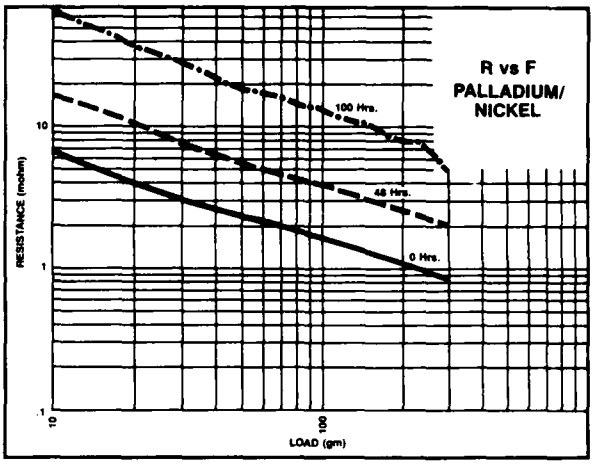


Fig. 15 - Contact resistance versus normal force for a palladium (over nickel) contact surface subjected to a test environment simulating an industrial exposure.

Corrosion of the palladium, due to the humidity and chloride content of the test environment are causing some degree of surface corrosion of the palladium. Finally, Figure 16, for palladium-nickel (80-20) alloy shows that this alloy experiences significant surface corrosion in the test environment.

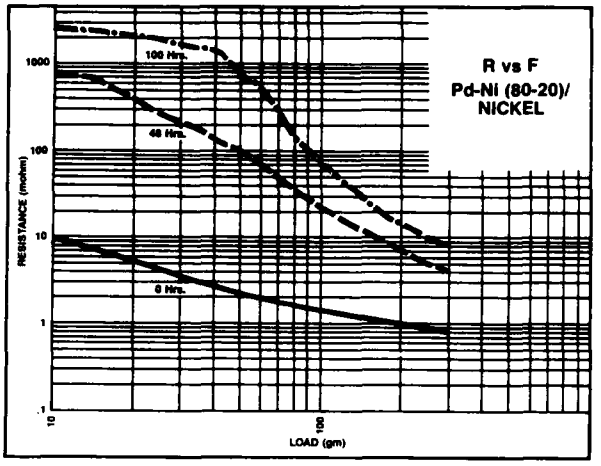


Fig. 16 - Contact resistance versus normal force for an 80/20 palladium/nickel (over nickel) contact surface subjected to a test environment simulating an industrial exposure.

The presence of the nickel, a non-noble metal, is the source of the reduced corrosion resistance.

These data show the uniqueness of gold as a contact surface in resisting environmental attack. It is important to note, however, that the effects of corrosion migration, indicated by the small contact resistance increases in Figure 13, must be considered even for gold contacts. More will be said on this subject in the discussion of connector housing functions.

NON-NOBLE FINISHES - This discussion will be limited to tin and silver since they are the most common non-noble contact finishes. Non-noble contact finishes are typicaly much thicker than precious metal finishes, 2.5 to 5 microns compared to 0.6 to 2. For this reason finish defects are minimal and the effects of underplates are less dramatic. Tin finishes are also used to provide enhanced solderability on the tails of contacts. Similar thicknesses apply in such applications.

Tin (and tin alloys) and silver will support surface films. Tin surfaces are always covered with an oxide film, while silver is subject to sulfide tarnish and chloride films. Tin oxide and silver sulfide films are readily disrupted on mating due to the relative softness of these materials. Silver films containing chlorides may be more resistant to disruption. Normal force requirements for non-noble finishes will be higher than for noble finishes to provide this disruptive capability. Wiping action on mating is also a requirement for non-noble contacts to facilitate displacement of surface films. In addition, normal force requirements will be higher to provide stability of the contact interface through the associated friction forces as mentioned earlier. Tin and silver have different performance characteristics and will be discussed separately.

As mentioned above, tin contact finishes are characterized by a surface oxide film. This oxide film must be disrupted for a reliable contact interface to be achieved. Figure 17 contains a schematic depiction of the disruption of a surface film on a soft metal, such as tin.

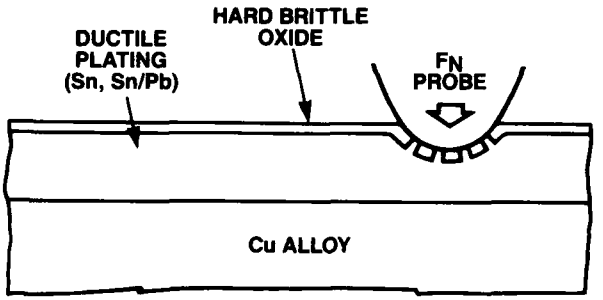


Fig. 17 - Schematic depiction of the mechanism of oxide displacement on a soft metal.

Application of a load to the typically thin and brittle tin oxide film results in cracking of the oxide and transfer of the load to the soft and ductile tin. The tin flows under the load resulting in separation of the cracked oxide and extrusion of the tin through the cracks. The end result of such a process, on aluminum, has been shown in Figure 3.

At this point a reliable contact interface has been established. The tin, however, remains susceptible to oxidation and this susceptibility leads to the major degradation mechanism in tin contact finishes, fretting corrosion.

The mechanism of fretting corrosion is schematically illustrated in Figure 18 (21). Low amplitude, hundredths to tenths of a micron, repetitive motions of the contact interface are referred to as fretting. Such motions can result in the tin contact interface being broken. If the interface reoxidizes during these disturbances the resultant effect is termed fretting corrosion. Fretting corrosion can lead to dramatic increases in contact resistance, including open circuits, as the oxide debris builds up at the contact interface (22). There are two approaches to avoiding fretting corrosion. First, contact lubricants, to protect the interface from corrosion, are used in some cases. Second, high normal forces, to produce correspondingly high friction forces which resist contact motion, are commonly used. Normal force requirements for tin contact finishes are higher than those for precious metal finishes for this reason.

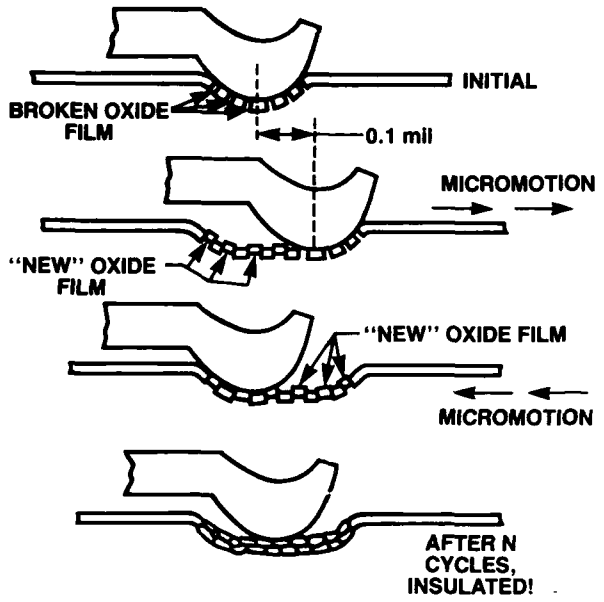


Fig. 18 - Schematic representation of the kinetics of fretting corrosion due to translational micromotions.

In contrast to the tens of grams recommended for gold, for example, a minimum of 100 grams is recommended for tin (23). Higher values of normal force may be necessary if the mechanical stability of the operating environment dictates, ie. high vibration applications. Application guidelines for tin contacts have been provided by Whitley (23).

Normal force requirements for silver have attributes of both noble and non-noble finishes. Silver is relatively corrosion resistant, but is susceptibe to sulfide tarnishing and chloride formation. The filming characteristics of silver are dependent on the application environment because of this combination of corrosion mechanisms. This is in contrast to tin where tin oxide provides a self passivating surface. Unlike tin, however, silver is not subject to fretting corrosion. Normal force requirements for silver are similar to those for tin.

Tarnish films on silver have resulted in performance degradation in low force non-wiping contacts in which surface films may not be completely displaced. The film can be semiconducting in such cases resulting in "diode" type behavior rather than the ohmic contact resistance desired of electronic contacts (10). The wiping action of typical separable connectors, however, is sufficient to deal with typical silver films. In some environments, however, silver films can be more difficult to displace. Another "limitation" to the use of silver results from its tendency for migration which can lead to "shorts" between contacts or contact pads. The importance of this limitation is very application sensitive, but must be a consideration when silver is a possible contact finish.

Silver contact finishes do have an advantage over other contact finishes in some respects. The electrical and thermal properties of silver are superior to any other contact finish. These characteristics result in silver providing superior performance as a contact for high current applications.

These general remarks on contact finish performance characteristics are relevant to any finishing method. The methods differ, however, in the general structure of the applied finish so that the characteristics of particular importance may vary with the finishing method.

FINISHING METHODS - There are several methods by which a contact finish can be applied. Three major technologies are:

1. Electrodeposition
2. Surface cladding
3. Hot dipping

In this discussion, the focus will be on the effects of each of these methods on connector performance, not on the techniques of the methods as such.

ELECTROPLATING - Electroplating, or electrodeposition, is the most widely used method of applying a contact finish to a contact spring. Numerous texts on electroplating practice and theory are available, see reference 25 for example, so the process will not be described in detail here.

BASIC COMPONENTS—GOLD PLATING SYSTEM

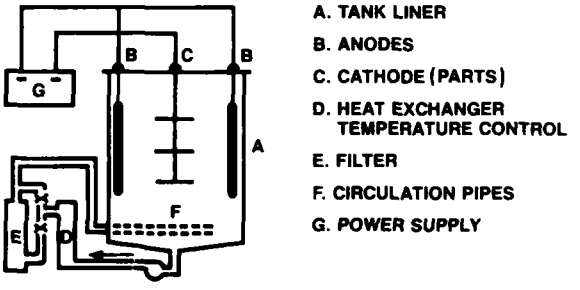


Fig. 19 - Schematic of a basic plating cell.

A typical electroplating cell is depicted in Figure 19. The plating is deposited on the cathode with the metallic ions coming from the plating bath, either through components of the bath or from a soluble anode which replenishes the bath as deposition occurs.

Current plating technology provides a number of options for selectively depositing the plating only in functional areas, in particular, the contact area. Development of such plating methods was driven by cost

considerations as the price of gold increased dramatically in the seventies.

In precious metal electroplates, porosity, base metal diffusion and corrosion migration, especially in selectively plated contacts, must be considered in connector design.

The relative importance and sensitivity of such considerations depends on the operating environment. The majority of connector operating environments are relatively benign. But as microprocessor technology expands, to process control applications for example, operating environments are becoming more severe and such corrosion considerations take on increasing importance. As mentioned above, nickel underplates have been successful in providing reliable performance of precious metal electroplates in most operating environments.

Gold and palladium, and alloys of these metals, are the most common precious metal electroplates. The hardness of electroplates, and hence the durability of the contact interfaces using them are dependent on the plating processes. Hardening agents such as cobalt and nickel can produce "hard golds" with hardnesses in the range of 130 to 240 Knoop. Palladium, and its alloys, provide even higher hardnesses, 150 to 500 Knoop. Composite platings of gold over palladium show exceptional durability performance as compared to gold or palladium alone (26). In these systems the gold appears to act as a solid state lubricant.

Tin, and tin-lead, electroplates are commonly used on electronic connectors. Tin-lead finishes have the advantage of not being subject to "whiskering", a filamentary single crystal growth morphology which can lead to shorting of contacts or contact pads. An example of tin whiskers is shown in Figure 20.

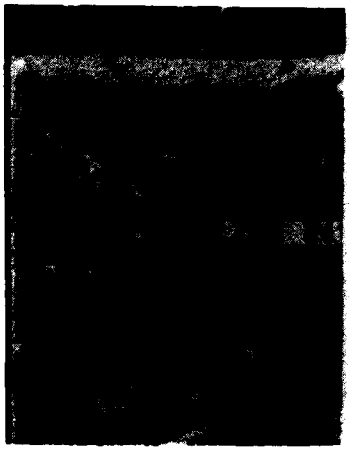


Fig. 20 - SEM photomicrograph of a tin contact surface showing tin whiskers.

The driving force for whiskering is the compressive stresses in the electroplate. This stress is reduced in tin-lead alloy plating. The lead content necessary to avoid whiskering is of the order of a few percent. Larger lead contents may compromise the environmental performance of the plating in corrosive environments. An alloy commonly used in connectors is 93 tin - 7 lead.

A major application of tin-lead electroplates in connectors is in providing a solderable coating. Solderability and solderable coatings will not be discussed in this paper, since many references are available (see 27 for example).

CLADDING (INLAYS) - Cladding consists of mechanically bonding metallic surfaces under high pressures (28). Inlays are a special case of cladding in which the material of the contact finish is selectively clad into the spring material as shown in Figure 21 (28). Skiving of the inlay groove provides a clean surface to promote bonding. The successive reductions used to get the strip metal to final gage produces the metallic bond required for the cladding process. Additional bonding occurs during the thermal treatments in processing.

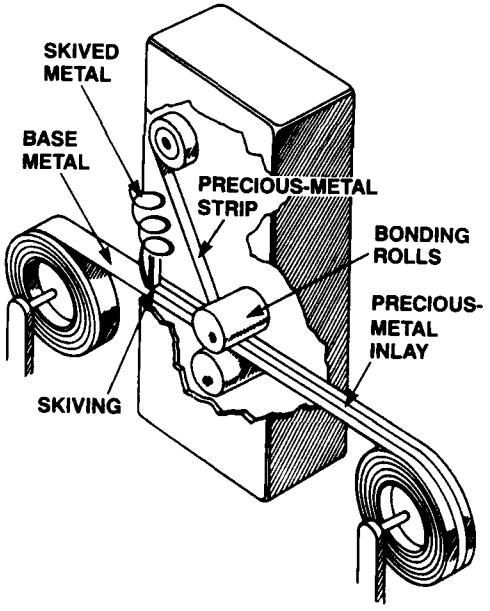


Fig. 21 - Schematic representation of an inlaying process.

Precious metal inlays include a nickel underplate to maintain the performance advantages of nickel as well as for the practical reason of providing a greater thickness of inlay strip to facilitate handling during processing.

Since inlays are wrought materials they are not subject to porosity in the same sense

as electroplates. Defects in the precious metal inlay can arise, however, and will have similar performance ramifications (29).

Cladding technology provides additional options in the choice of materials which can be used as contact finishes. Wrought precious metal alloys such as WE1 (gold, platinum, silver) and palladium-silver (60-40) are popular clad finishes which are not available as electroplates at this time. The palladium silver alloy has seen extensive use in telecommunications applications and is of interest as a precious metal finish option. Electroplating processes for the alloy are under development.

Tin and solder clads or inlays are also used in connectors, but not generally as contact interfaces. Such coatings are usually applied for soldering of the tails of the contact springs.

The combination of precious metal and tin or solder cladding on the same strip metal prior to stamping may offer manufacturing advantages which make cladding technology an attractive option as an alternative to electroplating.

HOT DIPPING - In connector applications hot dipping is limited to tin alloys. Hot dipping consists of pulling the strip metal through a molten bath to coat it with tin. Thickness control is accomplished by air knives or wipers. Typical thicknesses are, therefore, dependent on processing. Two general classes of thickness are typical, 1 to 6 microns using wipers, and upwards of 6 microns using air knives. Contacts are then stamped from the pre-tinned strip.

From a contact interface perspective, the most important difference between hot dipping and inlay or electroplated tin finishes is in the intermetallic compound that is formed in the hot dipping process. While copper-tin intermetallic compounds form, even at room temperature, in electroplated and clad finishes, the hot dipping process can produce significant amounts of intermetallic during processing if care is not exercised. Intermetallic compounds do not provide acceptable contact performance, and can also negatively affect the durability of contacts. The intermetallic thickness produced during hot dipping, therefore, must be carefully controlled to ensure that the contact surface is, in fact, tin and not intermetallic. It should also be mentioned that tin-lead and hot dipped tin finishes are not susceptible to whiskering.

Table I provides a summary of contact finish properties relevant to connector performance. The Table combines representative quantitative and comparative data. It is important to note that the range of properties can be even larger than listed in the table.

Table I - Summary of Contact Finish Considerations

FINISH	CONTACT RESISTANCE (mOhms @ 100g)	HARDNESS Knoop	COEFF. OF FRICTION	DURABILITY
Gold				
Cobalt	1.1	130/240	0.2/0.5	good/v. good
Inlay	0.6	25	1	fair/good
Palladium	1.4	150/350	0.3/0.5	good/v. good
Palladium/ silver 60/40(inlay)	1.7	100	0.7	v. good
Palladium/ nickel 80/20(plate)	8.0	400/500	0.3/0.5	v. good
Silver	0.5	50/125	0.5/0.8	fair
Tin				
Matte	0.8	10	0.6/1.0	poor/fair
Bright	1.0	20	0.4/0.6	poor/fair
Hot Dip*	1.2	10	0.3/0.6	poor/fair
Tin/Lead 93/7	0.7	10	0.5/0.8	poor/fair

* Values are very process dependent due to intermetallic compound effects.

Contact resistance values are given for finishes over phosphor bronze (C51100) substrates. Since connector durability is very design dependent, only comparative information is provided. These comparisons are based on a hemisphere to flat reference geometry. Selection of a contact finish depends on consideration of a wide range of criteria, both fundamental and application related. Application/environmental comments have been included in the text and are not repeated in the Table.

REQUIREMENTS FOR A CONTACT SPRING

The great majority of contact springs are made from copper alloy strip in relatively thin gauge, 0.1 to 0.8 millimeters. Manufacturing methods for these materials will not be discussed in this paper. A useful introductory treatment of this subject can be found in reference 30.

Contact springs perform two different functions in a connector:

1. To provide the normal force which produces and maintains the contact interface.
2. To carry the signal, or power, from the interface to the other unit to which it is connected.

The requirements for the second function are easily met by copper alloys. The conductivity range for alloys used in connectors ranges from 8 to 100 percent IACS. In signal applications the conductivity is of secondary importance in most cases. In such applications stability of the resistance is more important than the absolute value. For power applications, however, the conductivity can become a significant factor in material selection.

With respect to normal force, there are two different classes of requirements on a contact spring material. A typical contact will have two ends, one of which will generally be part of a separable interface, and the other, in most cases, a permanent termination. The material requirements will be similar for the two cases, but significant differences also exist. Let us consider the separable interface requirements first.

SEPARABLE INTERFACES - In the case of separable interfaces, the main function of a contact spring is to provide the required contact normal force. The material characteristics which are important in this regard are Young's Modulus and yield strength since these values strongly influence the deflection characteristics of the spring and the amount of deflection which can be supported. An additional requirement which must be satisfied is that the contact spring material have sufficient formability to be formed into the required contact shape under acceptable manufacturing conditions. The tradeoff between strength, due to normal force considerations, and formability, due to manufacturing requirements is a major factor in selection of contact spring material. A listing of pertinent materials characteristics for a few of the more common connector spring materials is provided in Table II. The tensile strength ranges listed are those typically used in connectors. For C17200, the values are for mill hardened material. The Table also includes a comment on comparative stress relaxation performance. Stress relaxation is important in that excessive stress relaxation of the spring during the application life of the connector may result in unacceptable losses in normal force.

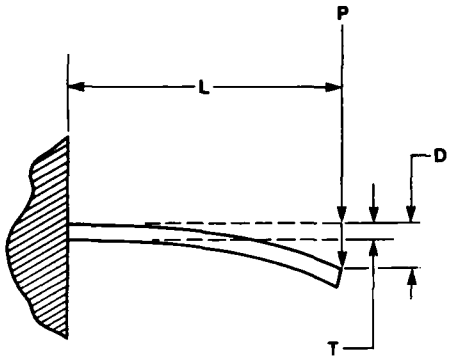
Using a simple cantilever beam as a contact spring let us consider how variations in these materials properties affect the contact normal force. For a cantilever beam, Figure 22, the force deflection equation takes the form

$$F = (D/4)\ E\ W(T/L)^3 \qquad (4)$$

where F is the force resulting from a beam deflection D, E is the Young's modulus of the spring material, and W, L and T represent the width, length and thickness of the beam respectively. Many contact spring configurations approximate this cantilever geometry.

Table II - Spring Characteristics

Material	Young's Modulus (10^9Pa)	0.2% Yield Strength (10^6Pa)	Stress Relax.	Form-ability	%IACS Cond.
C17200 (Be-Cu)	130	560-875	Excell.	Good-Excell.	21-26
C26000 (Brass)	112	420-630	Poor-Fair	Fair-Good	26-28
C51100) (Phos-bz)	112	490-770	Fair-Good	Good-Ver.Gd.	18-20
C72500 (CuNiSn)	140	490-665	Good-Ver.Gd.	Good	10-14



$$S = \frac{6PL}{WT^2} \quad \text{.....(1)}$$

$$D = \frac{4PL^3}{WT^3E} \quad \text{.....(2)}$$

S = MAX. STRESS; P = LOAD;
E = MODULUS OF ELASTICITY;
W = WIDTH; L.D.T. AS SHOWN

Fig. 22 - A cantilever beam with dimensions and equations relevant to calculation of contact normal force.

The only materials parameter in equation (4) is E, and it can be seen that the force for a given deflection of the beam increases as E increases. That is, the spring rate, given by F/D, of a given beam increases with E. For this reason, a high value of E is generally desirable since the normal force is larger for a given deflection. For the copper alloys most often used in connectors E falls in a relatively narrow range of 110 - 140 $x10^9$ Pascals. Young's modulus, therefore, is not a strong factor in materials selection.

A second equation which is useful in illustrating the effects of material properties on spring characteristics relates the maximum stress in the beam to its deflection.

$$Sm = (3/2)DET/L^2 \qquad (5)$$

where Sm is the maximum stress and the other variables are the same as in Equation (4).

Combining Equations (4) and (5) we obtain:

$$Fm = (1/6)\ SmWT^2/L \qquad (6)$$

This equation shows that the maximum normal force, Fm, which can be realized by a contact spring, while remaining elastic, is determined by the maximum stress the spring can support, which is the yield strength of the spring material.

Equations (4) and (6) show that the material characteristics of greatest impact in contact spring performance are Young's Modulus, which determines the spring rate, and the yield strength, which determines the maximum normal force which can be achieved under elastic loading of the beam.

Let us consider an example, taken from the literature (31) to demonstrate these interactions. The contact design, a post receptacle configuration, is shown in Figure 23.

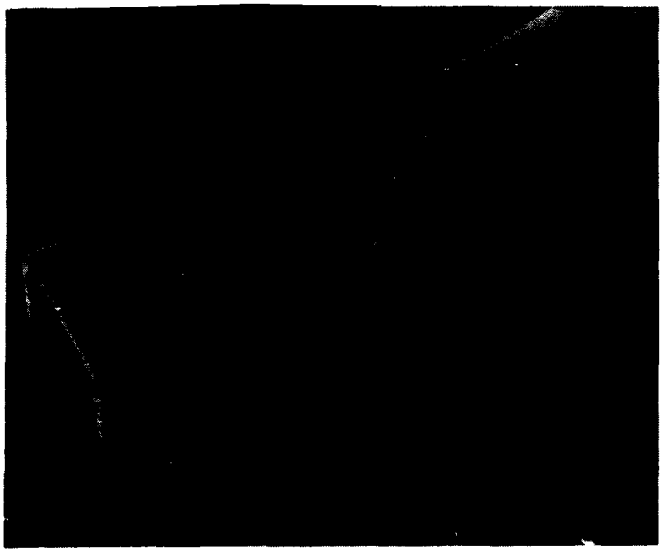


Fig. 23 - Optical photomicrograph of the receptacle contact used in generating the data in Figures 24 through 26. (Courtesy of Brush Wellman).

In this system the contact springs are essentially cantilever beams. Figures 24 through 26 are adapted, by elimination of some of the materials displayed, from reference 31. Figure 24 contains measured data on normal force versus beam deflection.

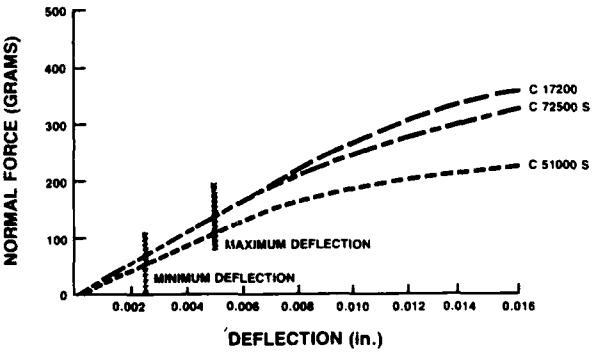


Fig. 24 - Relationship between normal force and contact beam deflection for the contact shown in Figure 23.

The data demonstrate the dependence of spring rate, the initial slope of the force-deflection curve, and maximum normal force on Young's modulus and yield strength respectively.

Table III - Material Properties for Example Contact

PROPERTY	C17200	C51000	C72500
Elastic Modulus (10^9 Pa)	126	118	134
0.2% off. Yield St. (10^6 Pa)	721	658	630
Elastic Limit (10^6 Pa)	336	315	182

From Table III, adapted from reference 31, we see that C17200 (beryllium-copper) and C72500 (copper-nickel-tin) have essentially equivalent values of Young's modulus which are slightly higher than that of C51000 (phosphor bronze). These modulus variations are reflected in the slopes in Figure 24 where the slope, which is the spring rate, F/D, of the C51000 is lower than the others.

Yield strength and elastic limit variations between the three materials are more striking as can be seen from Table III. These differences are reflected in Figure 24 by the deviation of the normal force versus deflection curves from linearity. The C72500 curve deviates first, followed by C51000 and finally C17200. The differences are more readily seen, however, in Figure 25 where permanent set versus deflection data are presented.

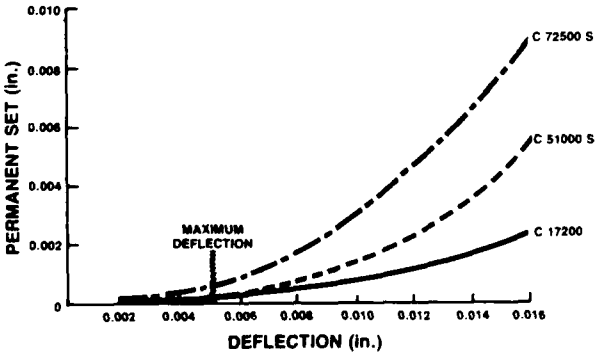


Fig. 25 - Relationship between permanent set and contact beam deflection for the contact shown in Figure 23.

The C72500 contact begins to take a set well before the C51000 and the C17200 contacts.

Permanent set can have significant effects on normal force since it reduces the available beam deflection. At the designed deflection of this contact, nominally 0.1 microns, the C72500 spring will take a set resulting in a loss in normal force on subsequent matings due to the reduction in available beam deflection. This effect is clearly seen in Figure 26.

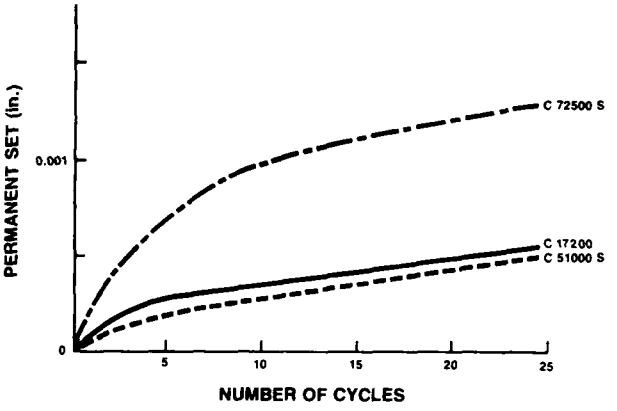


Fig. 26 - Beam permanent set versus the number of mating cycles for the contact shown in Figure 23.

After 25 mating cycles the C72500 spring has taken a set of the order of 0.025 microns while the C17200 and C51000 springs show no permanent set. The set experienced by the C72500 will result in a loss of normal force of about 25% since it represents one quarter of the deflection in this design. The significance of such a loss will depend on the magnitude of the normal force, the contact finish, and the application.

The impact of stress relaxation on normal force is straightforward. From equation (6) we see that normal force is directly proportional to the stress in the beam. Since stress relaxation is defined as a time dependent loss in stress under fixed deflection, which is the typical condition of operation of a contact spring, normal force will decrease with time at the same rate as stress relaxation occurs in the spring material.

From a mechanical viewpoint then, material selection criteria for a contact spring center about Young's modulus, the yield strength and stress relaxation resistance. Let us consider three of the more commonly used connector alloys, as listed in Table II, in terms of these criteria and an application/manufacturing perspective.

Brass, (C26000): Brass is the lowest cost contact spring material and is widely used in consumer connectors, primarily in high force systems. It provides good electrical conductivity and adequate strength for many applications. A major limitation of brass is poor stress relaxation behavior, which, in turn limits the temperature capability of contact springs made from brass. The formability of brass at its higher strength levels is also not suitable for many contact geometries.

Phosphor Bronze, (C51000, C51100 and C52100): In terms of the number of contact designs in which it is used, phosphor bronze may be the most common contact spring material. This high usage is related to the generally good materials characteristics of phosphor bronze alloys. The following remarks are general, and apply to varying degrees to the alloys listed. Phosphor bronze shows good strength and formability, and generally acceptable stress relaxation and conductivity. The good strength/formability performance of phosphor bronze allows manufacture of a wide range of contact geometries. The stress relaxation performance is good enough to allow its use over the normal temperature range, up to 105 C, usually specified for electronic connectors.

Copper-nickel-tin, (C72500): This alloy currently has limited use in connectors, being used primarily in telecommunications applications. Usage is limited by relatively low strength and electrical conductivity. These limitations outweigh the relatively good stress relaxation characteristics of the alloy.

Beryllium Copper, (C17200): Usage of beryllium copper is limited by the cost of the material. It provides superior strength/formability performance along with excellent stress relaxation characteristics. With these properties it is the material of choice for high temperature applications and will, most likely, see increasing use as connector miniaturization continues. As connectors decrease in size, or increase in density at the same size, the contact springs in the connector must become smaller. Reductions in the physical size of the contact will place increasing strength requirements on the contact springs since normal force requirements must still be met even at the reduced size. Beryllium copper is the copper alloy best suited to satisfy such strength requirements.

PERMANENT CONNECTIONS - Permanent connections are viewed as an extension of the conductor to which they are terminated. In that regard the resistance introduced by the connection is expected to be minimal, sometimes stated in terms of equivalent conductor length. In other words, the resistance introcuced by the permanent connection should be of the same order as a piece of the terminated conductor of the same length as the termination. Minimizing contact resistance, therefore, becomes a major design requirement of permanent connections.

There are several types of permanent connections which fall into several categories. One major separation is between soldered and mechanical connections. This paper will consider only mechanical connections. Within this category there are two basic classes, cold welded and residual force.

The first category, cold welded connections, are those in which the amount of deformation introduced during termination is such that a significant degree of cold welding occurs at the contacting interfaces. The many styles of crimps are in this category. For a more detailed discussion of crimp technology see Whitley (32).

During crimping the extensive deformation and relative motion of the conductor(s) against themselves and the terminal body results in a disruption of surface films producing new film free surfaces. Under the high forces of crimping, contact between these film free surfaces results in the formation of numerous microwelds. It is these cold welded joints which provide the electrical and mechanical integrity of crimps.

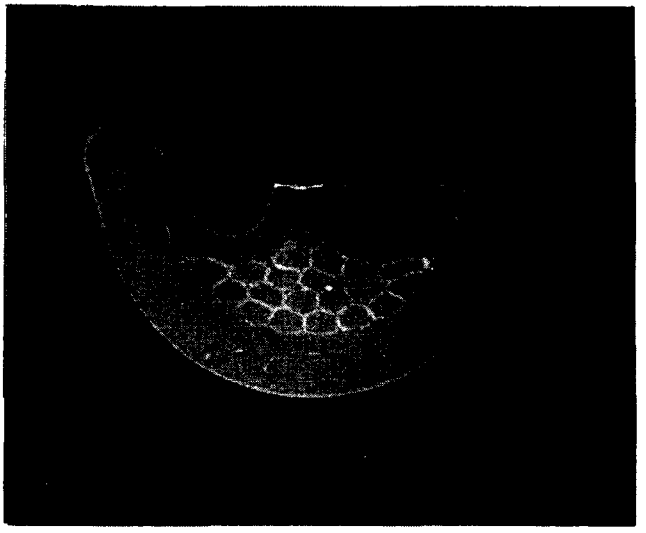


Fig. 27 - SEM optical photomicrograph of a cross section of a crimp. The extent of wire deformation is apparent (10X).

Figure 27 shows a cross section of a crimped connection. The deformation of the wire strands is evident. A schematic depiction of the change in some relevant engineering properties with deformation during the crimping process is provided in Figure 28. As deformation proceeds, the number of microwelds, and, therefore, the contact area, increases. As the number of microwelds increases the bonding between wires and the tensile strength of the crimp increases. As the contact area increases, the current capacity of the crimp and the electrical conductivity increase. The values peak as the interface contact area is optimized and begin to fall off as deformation reduces the cross sectional area of the conductor/crimp combination below that of the conductor. The deformation during crimping, therefore, is a parameter which must be controlled to ensure that the properties of the crimp remain in the "optimum" range. Because of the extensive deformation required in

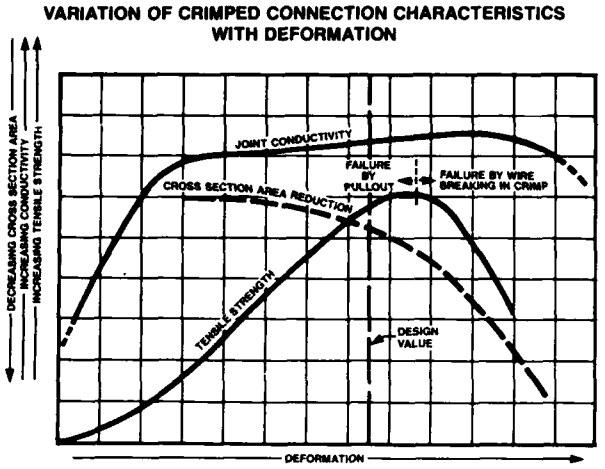


Fig. 28 - Schematic representation of the variation of crimp conductivity, tensile strength and cross section with crimp deformation.

crimping the ductility/formability of the terminal material is an important requirement which may limit the strength (temper) of the material with respect to spring characteristics. This limitation feeds back into the design/material requirements for the separable contact, which were discussed earlier.

The second category, residual force terminations, differs from crimping in the amount of deformation which occurs, being less than in crimping, and in the mechanism by which the contact interface is maintained. In residual force connections it is the elastic stored energy (springback) of the deformation which provides the normal force. For this reason it is the elastic deformation of the contact which contributes to normal force. There are two major types of residual force permanent connections currently in use, press in connections (using rigid or compliant pins), usually used in plated through holes on printed circuit boards, and insulation displacement (ID) connections, which are wire connections. Since they differ substantially, each will be discussed separately.

As mentioned, there are two types of press in connections, rigid and compliant. In both cases residual forces maintain the contact normal force. They differ in the relative amounts of elastic and plastic deformation which occurs in the pin and in the deformation produced in the printed circuit board.

Rigid pins produce more extensive board deformation and less contact deformation than compliant pins. Rigid pins, therefore, produce a lower residual force than compliant pins which, in turn, results in lower retention forces and reliability.

Many designs of compliant pins are in use, some of which are shown in Figure 29.

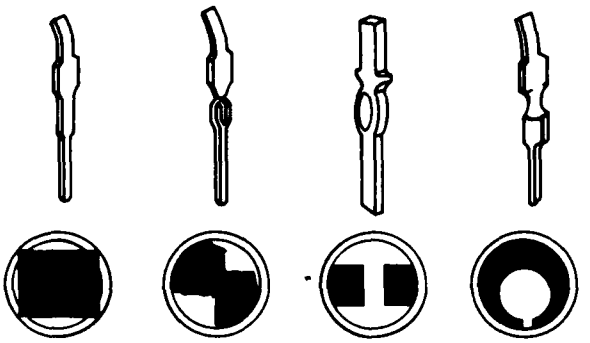


Fig. 29 - Schematic representations of compliant pin designs currently in use and cross sections through the pin when inserted into the plated through hole of a printed circuit board.

All are intended, in one way or another, to maximize the elastic deformation of the pin, and thereby the elastic stored energy, which maintains the contact normal force. Materials requirements for compliant pin designs depend on a tradeoff between ductility, to form the compliant section, and spring properties, to maximize the residual force.

Additional information on the design and function of compliant pins is provided in references 33 and 34.

There are also a wide variety of insulation displacement connection designs, some of which are shown in Figure 30 (35).

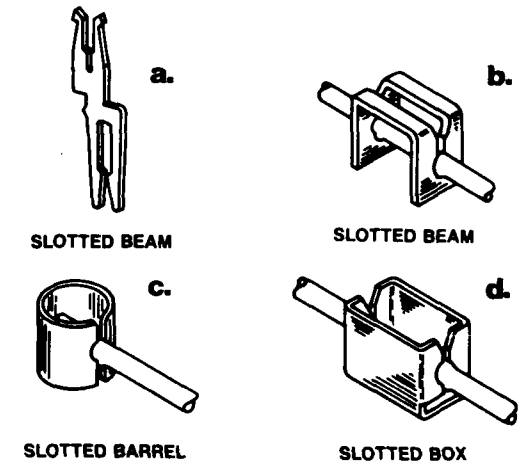


Fig. 30 - Schematic representations of IDC slot designs.

The simplest, which will be used to illustrate the principles of ID termination mechanics, is the slotted beam, Figure 30a. There are two separate functions which are critical to successful ID terminations:

1. the displacement, or stripping of the insulation.

2. the formation and maintenance of the contact area.

Figure 31 schematically depicts these two functions (36).

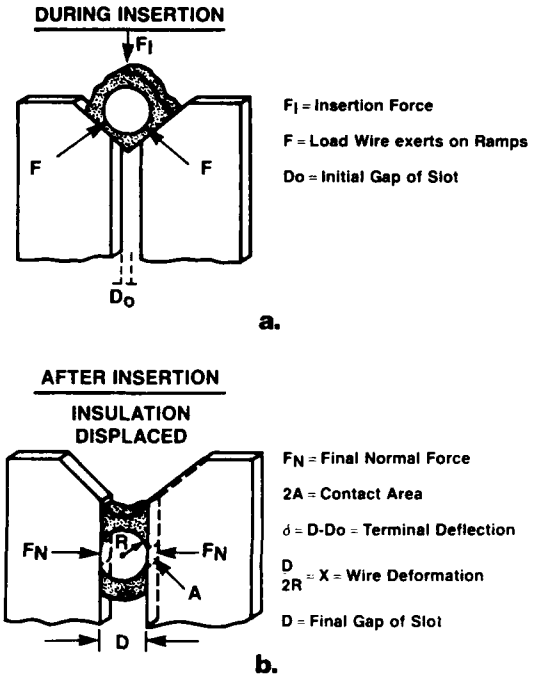


Fig. 31 - Schematic representations of the insulation displacement termination process. Figure 31a shows the forces on the conductor as it is inserted in the slot and the forces on the insulation which will lead to its displacement. Figure 31b shows the deformed wire which produces the contact area and the normal force which maintains it.

The beams of the terminal may undergo their maximum deflection during insertion of the conductor and displacement of the insulation, Figure 31a. In such cases, the maximum stresses on the beams, and, therefore, the requirements on the yield strength of the beam material are determined in this step.

After displacement of the insulation, the conductor continues downward in the slot to its final position, Figure 31b. It is important to note that extensive wiping action occurs during the downward motion producing a cleaning action on the beam sides and the conductor(s). At the final conductor location the displacement of the beams determines the residual normal force which maintains the integrity of the contact interface. The deformation of the conductor is limited by two factors. The lower limit is dependent on the requirement to establish a minimum contact area, A in Figure 31b. The upper limit of conductor deformation is due to the weakening of the wire and its effect on handling of the connectors during assembly and in its intended application.

The residual force in ID terminations is provided primarily by the contact beams. The conductor deformation is primarily plastic and contributes very little elastic stored energy. For this reason the materials characteristics of the contact beams are dominant in determining ID connection performance. There are two major requirements for ID connection materials. The first is a high yield strength so that acceptable elastic deformation of the beams, and, therefore, normal force, are achieved prior to plastic deformation. The second is that the beam material have acceptable stress relaxation characteristics.

Stress relaxation resistance is required since excessive stress relaxation may result in a reduction of the residual force to a level at which the integrity of the contact interface is compromised by mechanical disturbances, or by thermally induced stresses, during use.

To this point the materials requirements for residual force permanent connections are essentially the same as for separable connections. This is, in fact, the case with one modification. The normal force requirements for permanent connections are much higher than those for separable interfaces. This is to provide the minimum connection resistance and high reliability which is expected of permanent connections. Normal forces in separable connections are also lower due to the separability requirement. Mating forces and damage to the contact interface (durability) place a limit on the normal force which can be used in a separable connection.

In summary, selection of a material for a contact spring depends on a balance of application and manufacturing requirements. The materials characteristics of greatest importance are yield strength, formability, stress relaxation, and, of course, cost. The balance between these factors and the requirements of the application, especially temperature, establish the matrix from which the "appropriate" material should be selected.

For additional information on selection and characterization of contact spring materials see references 37 and 38.

REQUIREMENTS FOR A CONNECTOR HOUSING

The connector housing has three obvious, and one not so obvious - but very important - functions. The obvious functions are:

1. To electrically isolate the individual contacts from one another.
2. To maintain the dimensional spacing of the contacts.
3. To facilitate mating of the connector by providing mechanical protection and support for the contact springs.

The not so obvious function is to provide shielding of the contacts from the environment in which the connector is operating. In this way the susceptibility of a connector to corrosion can be reduced, but certainly not eliminated.

An illustration of the significance of connector housing shielding on connector performance is provided in Figure 32 (14).

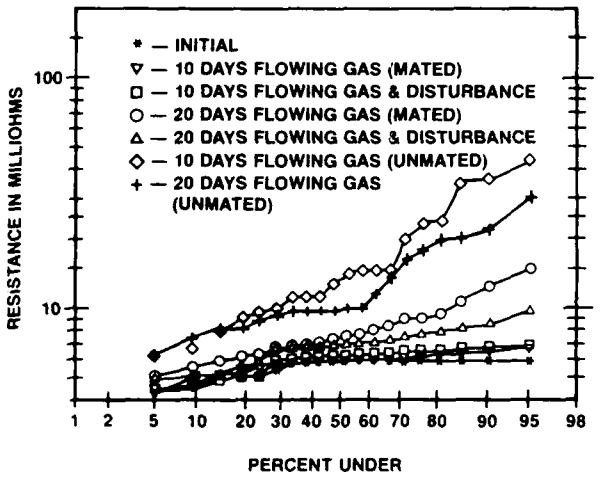


Fig. 32 - Contact resistance versus cumulative percent for separable connectors which have been exposed to a test environment intended to simulate an industrial environment. Data for both mated and unmated exposures are shown.

The data shown are for a post/receptacle connector, from the same product family as the design shown in Figure 1, exposed to a corrosive environment in both the mated and unmated condition. The environment and test exposure duration were intended to simulate a ten year exposure to an industrial environment (20). The data for unmated exposure shows the effect of the environment on the contact materials. Contact resistance has increased significantly after exposure. The data for the mated connector, on the other hand, show very good contact resistance stability. The housing has been very effective in shielding the contacts from the environment. This facet of connector design is an especially important consideration for connectors which are intended to operate in corrosive environments. Environmental shielding effectiveness, however, depends primarily on the connector housing geometry, rather than on materials properties, so we shall consider it no further in this discussion.

The other three functions, electrical isolation, dimensional stability and mechanical support, do depend on material properties. In fact, housing materials are selected primarily on the basis of the material properties which determine these functional characteristics.

It must be recognized that it is not only the value of a particular property which is important. The stability of the property with respect to manufacturing and application requirements also strongly influences material selection. In fact, it is on the basis of the stability of particular properties that most housing material selection depends.

The effectiveness of electrical isolation will depend on properties such as dielectric strength, and surface and volume resistivity. All the materials listed in Table IV will provide electrical isolation far in excess of that required in typical connector applications. They differ, however, in the stability of these characteristics with respect to exposure, particularly temperature and humidity. Because of the very high intrinsic values, however, such degradation may not affect connector performance.

The dimensional stability of a housing will depend on material selection and molding, assembly and exposure conditions. In this regard there are significant differences between materials as will be discussed.

Molding characteristics such as mold fill and shrinkage will determine whether a given connector design can be successfully molded with respect to meeting dimensional stability requirements. Such requirements increase in severity as connector requirements trend towards both miniaturization and higher pin counts. In both cases increased density of contacts becomes a requirement which impacts on the wall thickness between contact positions. The end result is that the moldability of thin sections becomes a key material selection criterion. Increasing connector length, to provide higher pin counts, introduces a need for the capability to fill long sections of housing external walls while maintaining the ability to fill the thin sections separating contact positions. Molding of connector housings is a demanding task for both housing and mold designer as well as for the materials and materials suppliers. An example of a high density miniaturized connector is shown in Figure 33. The product family for this connector comes in sizes as high as 684 positions in which case the connector is 47 centimeters long.

Fig. 33 - A collection of high density connectors. The number of positions varies from 60 to 636. Both plug and receptacle connectors are shown.

Assembly requirements, in particular soldering requirements, must also be considered. The trend towards surface mounting of connectors places particularly stringent temperature stability requirements on connector materials. One materials characteristic which can provide a guideline for the dimensional temperature stability of a material is the heat deflection temperature. Toughness, to withstand handling requirements, is also important. Snap fit and other assembly requirements may also impact material selection.

Meeting these requirements, however, is only part of the battle. Application exposure conditions, once again primarily temperature and humidity will affect material choice as well. Thermal expansion characteristics, especially expansion mismatch between the connector housing and the surface to which it is connected, another housing or a printed circuit board, for example, are particularly important. Moisture absorption, which is humidity dependent, may also be a consideration. The ambient temperature of the application environment is also a parameter of increasing concern. The magnitude of the concern will vary with the connector design, depending particularly on whether stress relaxation or creep of the polymer can significantly affect connector performance.

The mechanical strength required of a housing material is determined primarily by mating, "abuse" and contact spring support requirements. Mating of the connector requires some degree of toughness and mechanical strength, the magnitude of which depends on the connector design, in particular the number of positions. "Abuse" covers a variety of conditions including contact insertion on assembly of the connector, mismating during application and more conventional abuse such as mechanical shocks from a variety of sources. Toughness is a major requirement in providing "abuse" resistance. Connector housing design is also important in minimizing effects of abuse. Pre-engaging features and lead in ramps, for example, minimize the potential for abuse on mating. Contact spring support includes the retention of the spring in the housing, support of the spring in preload designs and, in some cases, mechanical support for contact normal force.

Selection of acceptable connector housing materials will involve tradeoffs based on application and manufacturing requirements and material characteristics in addition to those discussed here. Another important consideration is the approvals requirements of agencies such as UL, CSA and MIL.

Because of the wide variety of polymers available, only a few of them which are commonly used will be discussed, with apologies to the manufacturers of those which are omitted. There is, of course, an extensive literature on polymer properties and characteristics and only a few references are offered here (39-45).

Before specific materials are considered, a brief discussion of polymer materials in general may be appropriate. There are many ways to organize such a discussion, and our focus will be on materials characteristics as they affect connector performance.

For our purposes there are two classes of polymer materials, thermosets and thermoplastics. Thermosets are "permanently" set when they are molded. Thermoplastics are capable of reprocessing, that is, reject parts may be suitable for reuse. This economic advantage of thermoplastics, in combination with the greater range of materials and material properties available have resulted in their domination of connector applications. Thermosets are used primarily in military and high temperature applications and will not be discussed further.

In the thermoplastic polymers, there are two basic classes, crystalline, in which the polymer chains are aligned to a significant degree, and amorphous, in which the chains are randomly oriented. As would be expected, this structural difference affects many performance characteristics.

In general, crystalline polymers show greater anisotropy in properties and in molding characteristics such as warpage, better chemical stability and resistance to solvents, and superior mechanical characteristics.

Amorphous polymers are isotropic and, therefore, show superior molding characteristics with respect to mold shrinkage and warpage. These positive characteristics

are counterbalanced by lower strength, temperature stability and, frequently, poor chemical/solvent resistance.

Both amorphous and crystalline polymers are often, or generally, strengthened by the addition of reinforcing agents, predominantly glass fibers. Reinforcement increases the mechanical strength and temperature capabilities of the materials, and reduces mold shrinkage. Negative characteristics of reinforcement include increased mold wear and difficulty in molding, as well as an increase in anisotropy, particularly with respect to mold warpage and the capability to fill thin sections. On balance, however, the positive effects on mechanical strength and stability support the use of glass reinforcements in connector applications.

With this limited background, let us consider some of the more commonly used connector housing materials.

POLYAMIDES (NYLON) - At one time 6/6 nylon was a dominant connector housing material. This nylon possesses a good balance of mechanical, electrical and temperature characteristics, good chemical resistance and stability and good processibililty. It has, however, one severe limitation, a tendency to absorb moisture. Moisture absorbtion has a negative effect on the mechanical properties of nylon, and a significant negative effect on its dimensional and electrical stability. This limitation, in combination with the development of a wide variety of alternative materials, has diminished the range of applications for nylon as a connector material.

POLYBUTYLENE TEREPTHALATE (PBT) - PBT is arguably the most commonly used connector housing material due to its combination of processibility and functional characteristics. PBT is a crystalline material and possesses good electrical and mechanical properties, dimensional and chemical stability, good solvent resistance, and generally acceptable temperature capability. Temperature capability, however, does limit the acceptability of PBT in surface mounting applications. The heat deflection temperature (HDT) of PBT, 204°C, is too low for vapor phase soldering at 215°C, a commonly used surface mounting method. PBT is generally used in reinforced grades for improved mechanical properties.

The molding characteristics of PBT are, in general, good. The resin must be properly dried prior to molding, however. The ability of PBT to fill thin sections, particularly in reinforced grades, is limited and some degree of warpage is experienced.

POLYETHYLENE TEREPTHALLATE (PET) - PET shares the structural and performance characteristics of PBT with an improvement in temperature capability, an HDT of 224°C. For this reason, PET is one of the materials of choice for surface mount connectors. On the negative side, the creep and warpage of PET are greater than PBT. Creep under load is a concern in connectors in which the contact spring is preloaded by the housing. Once again, glass reinforcement improves the mechanical properties. PET resins are generally reinforced with 30 to 50 percent glass. Drying of the resin prior to molding is also critical to proper processing of PET.

POLYPHENYLENE SULFIDE (PPS) - PPS is also crystalline. It has high temperature capabilities, HDT 260°C and a continuous use temperature of 210°C. The mechanical properties of PPS are also very good, particularly with respect to stiffness. The negative side of this strength, however, is a tendency towards brittleness.

The molding characteristics of PPS are generally good, with an ability to fill thin sections and long cavities, and a minimum degree of warpage. A negative molding feature is a tendency to flash, extrude thin layers of material in the gaps between mold sections. The flash must be removed in a secondary operation.

POLYETHERIMIDES (PEI) - PEI is an amorphous material and as such shows desirable molding characteristics such as low warpage, even in non uniform sections, and low shrinkage. It also shows good elongation or toughness although with limited mechanical strength. Mechanically, PEI has limited creep resistance and a low modulus. The temperature capability is also low, HDT 210°C. PEI copolymer resins have been developed with significantly improved temperature capabilities. As an amorphous material, the solvent resistance of PEI is also limited. PEI is susceptible to moisture absorption and must be dried prior to molding.

LIQUID CRYSTAL POLYMERS (LCP) - LCP are a relatively new materials characterized by high stiffness and mechanical strength. These mechanical characteristics are maintained at elevated temperatures. The mold flow characteristics of LCP are very good with a capability to fill thin wall sections and near zero shrinkage. LCP is a good candidate for surface mount applications since its thermal expansion characteristics approach those of epoxy printed circuit boards.

Representative values of some of the relevant properties for the connector housing materials discussed are listed in Table IV. The values are for glass reinforced grades in the range of 25 to 40 percent glass.

Table IV - Housing Material Properties

	PA	PBT	PET	PPS	PEI	LCP
DIELECTRIC STRENGTH(KV/mm)	18	24	26	18	31	44
TENSILE(10^6Pa) STRENGTH	133	119	154	126	168	168
FLEXURAL MODULUS(10^9Pa)	7.0	7.7	10.5	11.9	8.4	14.7
NOTCHED IZOD(Joules/M) 75 C	54.9	37.6	46.2	37.6	57.8	63.6
HDT($1.8x10^6$Pa,°C)	241	204	224	260	210	240
T-INDEX(°C)	130	135	150	210	200	130

The process of selection of a polymer for a connector housing is complicated by the variety of materials available and the rapid rate of introduction of new materials. The electrical characteristics of engineering polymers are far in excess of typical connector requirements in most cases. Strength and dimensional requirements and the particular requirements of an application are the factors most likely to influence material selection. In general, more than one material will be acceptable, and selection is often influenced by experience and familiarity as well as technical characteristics.

CONCLUSION

This paper has presented a review of materials requirements for electronic connectors. Materials, process and application characteristics and requirements, and their interactions, have been discussed. Electronic connectors are truly a materials system in which interactions of the many functions and materials involved can significantly affect the performance and reliability of a connector design.

ACKNOWLEDGEMENT

Thanks are due to many co-workers at AMP for their help and comments on this paper. Special thanks go to Barb Siler for her efforts in producing the final copy.

REFERENCES

1. J. H. Whitley, "Anatomy of a Contact", Insulation/Circuits, October, November, December, 1981

2. J. B. P. Williamson, "The Microworld of the Contact Spot", Proc. 27th Annual Holm Conference on Electrical Contacts, Chicago IL, Sept. 21-23, 1981

3. J. R. Osias and J. H. Tripp, "Mechanical Disruption of Surface Films on Metals", Wear, Vol. 9, pp. 388-397, 1966

4. I. H. Brockman, C. S. Sieber and R. S. Mroczkowski, "A Limited Study of the Effects of Contact Normal Force, Contact Geometry and Wipe Distance on the Contact Resistance of Gold Plated Contacts", Proc. 38th Elec. Comp. Conf., Los Angeles, CA, May 9-11, 1988

5. I. H. Brockman, C. S. Sieber and R. S. Mroczkowski, "The Effects of the Interaction of Normal Force and Wipe Distance on Contact Resistance in Precious Metal Plated Contacts", Proc. 34th IEEE Holm Conference on Electrical Contacts, Chicago, IL, Sept. 26-29, 1988

6. D. Tabor, "Friction - the Present State of Our Understanding", Trans. ASME Jnl. Lub. Tech., Vol 103, pp. 169-179, 1981

7. J. A. Greenwood, "Constriction Resistance and the Real Area of Contact", Brit. J. Appl. Phys., Vol. 17, pp. 1621-1632, 1966

8. F. L. Jones, The Physics of Electrical Contacts, Oxford University Press, London, England, 1957

9. R. Holm, Electric Contacts, Springer Verlag, New York, NY, 1967

10. H. N. Wagar, Chapter 8, Principles of Electronic Contacts, in Physical Design of Electronic Systems, Vol. III, Prentice Hall, Englewood Cliffs, NJ, 1971

11. J. H. Whitley and R. S. Mroczkowski, "Concerning Normal Force Requirements for Gold Plated Contacts", Proc. 20th Annual Connectors and Interconnection Technology Symposium, Philadelphia, PA, Oct. 19-21, 1987

12. R. H. van Horn, "The Design of Separable Connectors", Proc. 20th Elec. Comp. Conf., May 13-15, 1970

13. M. Antler, "Wear of Contact Finishes: Mechanisms, Modelling, and Recent Studies of the Importance of Topography, Underplate and Lubricants", Proc. 11th Annual Connectors and Interconnection Technology Symposium, Cherry Hill, NJ, Oct. 25-26, 1978

14. R. S. Mroczkowski, "Corrosion and Electrical Contact Interfaces", Presented at NACE Corrosion 85, Boston MA., March 25-29, 1985

15. R. H. Zimmerman, "Metallurgical and Chemical Engineering of a Pressure Connection", presented at EDN Regional Engineers Meeting, New York, NY, December 3, 1964

16. R. S. Mroczkowski, "Precious Metals as Contact Platings", Proc. Ninth International Prec. Met. Inst. Conf., New York, NY, 1985

17. R. Cobaugh and A. S. Taylor, "Card-Edge Connector Design for Cost Effective Backplanes", Proc 13th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, Oct. 7-9, 1980

18. R. H. Zimmerman, "Plating, Diffusion, and Oxide Films on Electrical Contacts", Invitational Symposium, Montreal, Canada, 1963

19. M. Antler and M. H. Drozdowicz, "Wear of Gold Electrodeposits: Effect of Substrate and of Nickel Underplate", Proc 24th Annual Holm Conf. on Electrical Contacts, Chicago, IL, Sept. 11-15, 1978

20. W. H. Abbott, Private Communication, Battelle Laboratories, Columbus, OH

21. R. S. Mroczkowski, "Connector Contact Surfaces - Where the Action Is", Presented at Indicon, Indianapolis, IN, August 23-24, 1983

22. E. M. Bock and J. H. Whitley, "Fretting Corrosion in Electric Contacts", Proc. 20th Annual Holm Conference on Electric Contacts, Chicago, IL, Oct. 29-31, 1974

23. J. H. Whitley, "The Tin Commandments", Plating and Surface Finishing, October, 1981

24. H. N. Wagar, Ref. (10), Chapter 9

25. L. J. Durney (Ed.), Electroplating Engineering Handbook, Van Nostrand Reinhard Co., New York, NY, 1984

26. T. Sato, Y. Matsui, M. Okada, K. Murakawa and Z. Henmi, "Palladium With a Thin Gold Layer as a Contact Material", Proc. 26th Annual Holm Conference on Electric Contacts, Chicago, IL., Sept. 29 -Oct. 1, 1980

27. C. A. MacKay, "Surface Finishes and Their Solderability", Welding and Metal Fabrication, Jan./Feb. 1979

28. C. Harlan, "Overview: Inlay Clad Metal and Other Electrical Contact Materials", Metals Progress, November, 1979

29. R. J. Russell, "Properties of Inlay Clad Wrought Gold Alloys", Proc. 7th Annual Connector and Interconnection Symposium, Cherry Hill, NJ, Oct. 23-24, 1974

30. J. H. Mendenhall Ed., Understanding Copper Alloys, Winchester Press, NY, for Olin Corporation, 1977

31. W. S. Loewenthal, J. C. Harkess and W R. Cribb, "Performance Comparison in Low Deflection Contacts", Presented at Internepcon/UK84, 1984

32. J. H. Whitley, "The Mechanics of Pressure Connections", Presented at EDN Regional Engineers Meeting, New York, NY, December 3, 1964

33. R. Cobaugh and F. Miller, "Compliant Pin - An Idea Whose Time Has Come!", Presented at the IPC Meeting, Washington, DC, 1978

34. F. J. Dance, "The Electrical and Reliability Characteristics of Compliant Press-fit Pin Connectors", IPC Technical Review, December 1987

35. R. S. Mroczkowski, "Insulation Displacement Technology" AMP Engineering Note EN 133, 1985. See also, Connection Technology, July, 1986

36. R. D. Malucci, "A Unique Method To Measure The Loading Characteristics Of Insulated Wire As It Is Inserted Into A Slot", Presented at the Ninth International Conference on Electric Contact Phenomena, Chicago, IL, Sept. 11-15, 1978

37. M. J. White, "Interrelationships Between Contact Alloy Properties Relative to Connector Design, Manufacture and Performance", Proc.35th Elec. Comp. Conf., Washington, DC, May 20-22, 1985

38. T. E. Bersett, "Back to Basics: Properties of Copper Alloy Strip for Contacts and Terminals", Proc. 14th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, 1981

39. L. J. Dean, "Connectors: The Challenge Gets Tougher", Plastics Design Forum, July/August 1983

40. J. F. Zappala and D. Blumenfeld, "A Systematic Approach to Selecting High Performance, Cost Effective Connector Insulating Resins", Proc. 15th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, Nov. 1-2, 1982

41. R. Walczak and P. Richards, "Designer's Guide for Use of Thermoplastic Polyesters in Today's Connectors", Proc. 20th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, Oct. 19-21, 1987

42. R. E. Wright, V. R. Landi and M. J. DeLassus, "Connector Body Insulation", Proc. 16th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, Nov. 11-14, 1983

43. M. H. Horn and P. R. Kemper, "Thermoplastics for Connector Bodies", Proc. Third Annual ASM Conf. on Elec. Packaging, Minneapolis, MN, April 28-30, 1987

44. A. Tolmie, "Liquid Crystal Polymers-The Thermoplastic Solution for Innovative Connector Designs", Proc. 20th Annual Connector and Interconnection Technology Symposium, Philadelphia, PA, 1987

45. E. Galli, "Connectors: A Continuing Drama in Plastics", Plastic Design Forum, May/June 1987.

ALLOY REQUIREMENTS FOR ELECTRONIC CONNECTORS

John F. Breedis
Olin Corporation
Metals Research Laboratories
New Haven, Connecticut, USA

ABSTRACT

A connector's performance is controlled by its design and the mechanical and physical properties of the alloy used in its construction. Important alloy properties include those which affect contact force, such as stiffness in bending and resistance to relaxation under stress while in service. Copper alloys are usually chosen for use in connectors because of the spring properties they afford, combined with good electrical/thermal conductivity, high corrosion resistance and capability for being soldered and electrolytically plated. This review is concerned with those aspects of copper alloy composition and processing that influence contact force at ambient and elevated temperatures. Correct application of alloy test data can predict connector behavior. In addition to the requirement for stability of contact force, the trend towards more densely spaced electronic components requires higher strength materials which can be made into more compact designs where alloy strip must be formed over smaller radii. The mechanical metallurgy of bend formability is discussed as it pertains to fabrication of connectors from copper alloys suited for this application.

INTRODUCTION

The increasing sophistication of electronic systems requires that connectors used in these systems provide better stability. Smaller components, packed more closely, operating under higher ambient temperatures (an extreme being under-the-hood auto applications) place added importance upon the selection of alloys used for the contact spring member of connectors. The ultimate choice of material for a particular connector is a balance of factors involving function, fabrication, and cost.

Connector reliability is directly related to the force maintained between contacts, assuming that contact resistance and plating and plating stability are not being affected by the environment. The initial contact force is dependent upon the load deflection properties in bending of the alloy being used. The extent to which this force can be maintained during service is a consequence of the alloy's resistance to stress relaxation at room and elevated temperatures. Stress relaxation caused by increased temperature from ohmic heating is generally not important for electronic connectors where the currents carried are low and alloys having lower electrical conductivities may suffice. This may not hold for high power applications.

Table 1

SELECTED COPPER CONNECTOR ALLOYS

UNS Designation	Nominal Composition (Pct)	Temper*	Mechanical Properties 0.2% YS (ksi)	TS (ksi)	Elong. (Pct)
C151	0.1 Zr	HR04	55	58	4
C172	1.9 Be	TM02	110	127	12
C1741	0.3Be-0.45Co	HT	105	118	5
C510	5Sn-0.1P	H08	100	102	2
		HR08	90	98	9
C521	8Sn-0.1P	H02	65	80	35
		H08	106	111	6
C7025	3Ni-0.75Si-0.1Mg	TM00	90	101	7
C724	12Ni-2Al-0.1Mg	TM02	110	135	12
C725	9.5Ni-2.3Sn	H02	69	72	5
		H04	80	83	2
		H08	90	93	

*HOX : Cold Rolled
HROX: Cold Rolled + Stabilized
TMOX: Mill Precipitation Hardened

The mechanical metallurgy of copper-base alloys is reviewed in the following relative to their performance in connectors and their fabrication. Performance is dependent upon metallurgical factors that effect spring stiffness and stress relaxation. Formability in bending, the principal mode used for connector components, is also discussed to provide a background against which material selections can be made. The alloys selected for illustration in this review are listed in Table 1.

PROPERTIES AND CONNECTOR PERFORMANCE

Load Deflection Response

Connectors are usually designed as springs which operate in bending. Deflection of the spring contact is a function of applied load, the dimensions and design geometry of the spring member, and the modulus of the alloy. Application of finite analysis techniques makes it possible to predict normal forces for complex connector geometries. However, the modulus term has not been chosen with such sophistication.

Beam formulae typically assume constant modulus, and use Young's modulus, determined in tension, for both low and high stress applications. However, most connectors operate in bending to strains beyond the proportional limit where use of an elastic modulus is not appropriate.

A better approach to defining a proper modulus is illustrated by Figure 1 (References 1-4). The elastic modulus is defined by the slope of the stress-strain curve, here shown in bending, at zero stress. Also shown is the secant modulus which is defined by slope of

the line connecting the origin with the design stress. Since the normal force at a contact is determined by the stress acting at a given deflection (strain) in a connector, it is apparent that using elasticity to predict force can result in a significant overestimation, the magnitude of which depends upon the designed deflection.

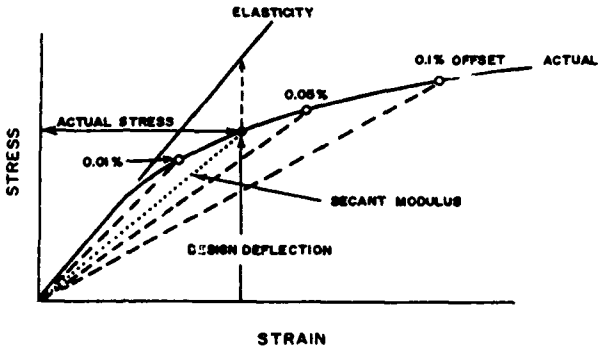


Figure 1: Stress-strain curve in bending. Shown is the secant modulus which can used to more correctly calculate contact force than the elastic modulus in bending at deflections that are typical for actual connectors.

It is also important to recognize that the secant modulus of copper alloys can be modified strongly by metallurgical factors related to processing, orientation relative to the strip's rolling direction, and alloy composition. The elastic modulus is much less dependent upon these factors.

Load-deflection response can be depicted as shown in Figure 2 where the secant modulus is plotted against applied, or design, stress. The typical appearance of such curves show no dependence of modulus upon stress initially, followed by a decrease in modulus as the outer fiber stress approaches 0.01% offset and beyond. Materials which exhibit higher moduli for a given stress will develop a greater normal force in a given design. In addition to providing parameters for connector design, such data presentation enables alloys and different temper conditions for a given alloy to be compared.

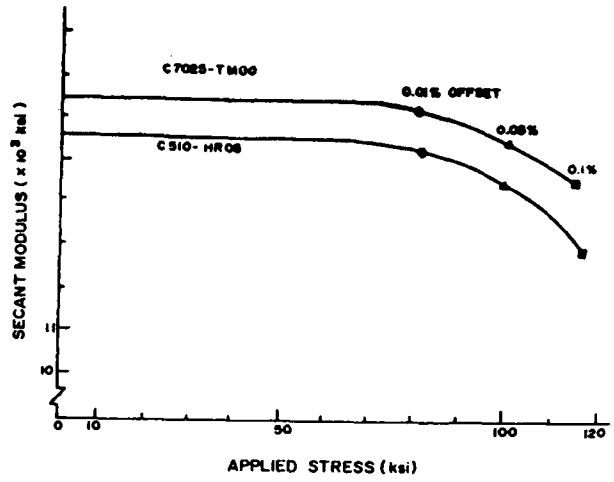


Figure 2: The dependence of secant modulus upon applied stress for spring temper C510, which has been stabilized, and mill hardened C7025.

Figure 3 depicts how load-deflection response can vary with temper, as well as orientation of the spring's length relative to the rolling direction of the strip from which it was fabricated.

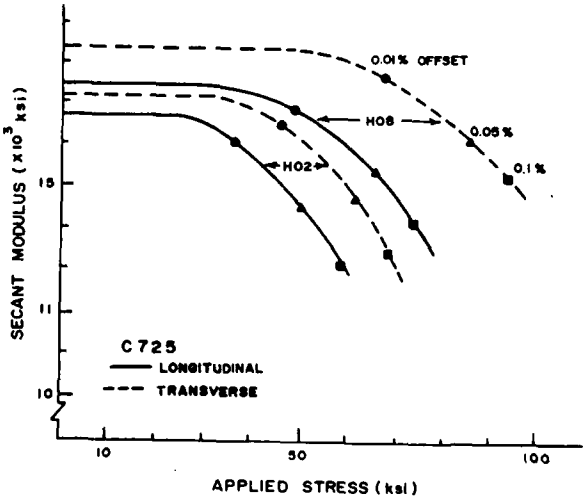


Figure 3: Temper and orientation dependence of load deflection response, as illustrated by C725 at two strength levels (69 and 90 ksi yield strengths) and in two strip orientations: longitudinal where the axis of bending for the spring is normal to the strip's rolling direction, and transverse where this axis parallels the rolling direction.

Directional effects increase with strength in cold rolled-to-temper alloys. The extent to which this directionality occurs is also alloy dependent. Further, low temperature heat treatment of rolled temper copper alloys (referred to as stabilization) generally improves stiffness and reduces anisotropy in load/deflection response. Alloys which are provided in mill hardened conditions, such as the precipitation hardened beryllium coppers: C172 and C1741, as well as the nickel-aluminide hardened alloy: C724, typically provide high stiffness in connector designs.

Normal Force Stability

The in-service reliability of a connector system depends upon maintaining contact force while under applied stress, often during elevated temperature exposure in the 75-100°C (167-212°F) range and in some cases higher, such as 125-150°C (257-302°F) for some automotive applications and 200°C (392°F) for electronic burn-in sockets. Stress relaxation is the time dependent decrease in contact force (stress) occurring at a fixed displacement (strain). This phenomenon is allied to creep which refers to a change in displacement occurring under a fixed load. Both phenomena are thermally activated and increase in rate with increasing temperature.

For stress relaxation in connectors, where displacement is invariant after the components are mated, dislocation motion occurs to relieve some of the initially imposed strain. The resulting decrease in elastic strain appears as a loss in contact force. Microstructural changes during stress relaxation depend upon alloying and the microstructure developed by processing to temper.

Stress relaxation behavior of an alloy is determined from tests in bending, which is consistent with the end application, and with samples that are tapered in width to maintain constant outer fiber stress over the test gage length. Initial stressing can range from 50 to near 100% of the 0.2% yield strength, measured in tension, without significantly affecting the outcome (Reference 5). The actual load remaining during testing can be measured directly, converted to stress, and then plotted as a function of time at temperature. When linear, the data can be extrapolated to predict the stress at the contact in actual, more prolonged service.

Examples of such data are presented in Figure 4 for the cold rolled to temper phosphor bronze alloys: C510 (5%Sn) and C521 (8%Sn).

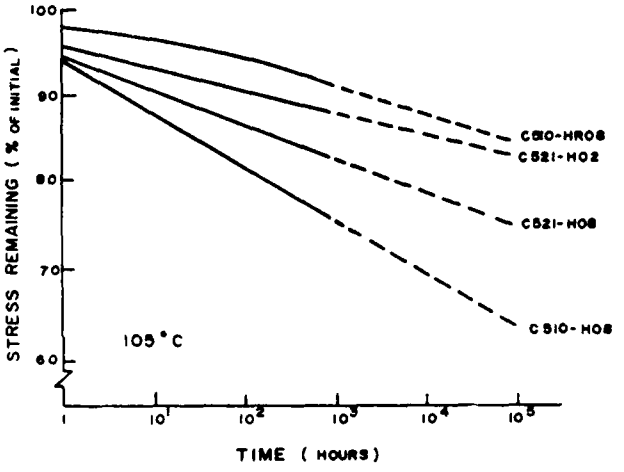


Figure 4: Stress relaxation behavior at 105°C (221°F), extrapolated to 10 years' (10^5 hours) service. Shown is the percentage of the initially imposed stress (up to 0.2% yield strength) for phosphor bronze alloys before and after stabilization (longitudinal orientation).

These curves show two effects: (1) the percent stress remaining decreases with increasing temper and (2), a low temperature thermal treatment (near 200° which decreases yield strength slightly) significantly improves stress relaxation resistance.

Copper alloys, all processed to about the same strength can differ in their resistance to stress relaxation, as depicted in Figure 5 (References 5-9). Here, rolled temper C725-H04, stabilized C510-HR08, and the precipitation (mill) hardened alloys: C7025 and C1741,

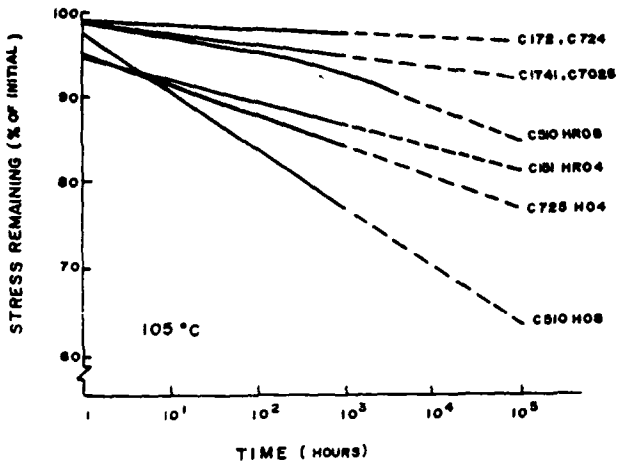


Figure 5: Comparison of stress relaxation behavior at 105°C (221°F) for selected copper connector alloys (longitudinal orientation).

are compared after processing to around 90-100 ksi yield strength. C172 and C724 are also represented in their mill hardened TM02 (110 ksi yield strength) temper; percentage stress remaining of the latter remains above 90% for the HM (130 ksi nominal yield strength) temper. Unusual among these alloys is C151, a relatively dilute alloy having high conductivity but lower strength. In the stabilized condition, the fine dispersion of Cu_3Zr precipitate and retained Zr in solution (near 0.01%) contribute to respectable stress relaxation resistance for use in those applications where ohmic heating must be minimized.

Different levels of stress relaxation resistance between alloys such as C510 and C725 can be rationalized, at least in qualitative terms, by atom misfit in those solid solution alloys and the amount of solute in solution (Reference 10). The atomic size difference with copper is greater for tin over nickel, and C510 has more tin in solution. The incremental effects of solute additions are greatest in dilute alloys, and become weaker at high solute levels. For example, up to 2-3% tin very strongly improves both resistance to stress relaxation and strength, with less improvement thereon. Cold work is also a factor. The amount of cold work required for reaching a given strength is less for C521 (8%Sn) versus C510 (5%Sn), for instance. Accordingly, C521 can provide higher stress relaxation resistance than C510, when both are at the same strength, through the combined effects of higher tin content and less cold work in the former alloy.

Low temperature heat treatment increases resistance to stress relaxation by promoting dislocation rearrangement into configurations which have fewer potentially mobile dislocations. Mill hardened alloys are typically heat treated as the final operation, usually at higher temperatures than used for stabilization of solution hardened, cold rolled alloys. The high stress relaxation resistance of precipitation hardened alloys result from the stabilizing effects of fine precipitate distribution, combined with dislocation rearrangement during heat treatment and solid solution effects. Typically, the mill hardened alloys can provide the highest stress relaxation resistance available from copper alloys, along with being very high strength materials.

Conductivity

As a general trend, copper alloys having higher strengths achieve this property at some sacrifice in electrical, and accordingly thermal, conductivity. This trend is illustrated in Table 2. Electrical conductivity can be important for some electronic applications where voltage drops are rationed because of available total power.

Where the currents being carried are low, and connector body cross sections are large relative to wires and printed circuit patterns carrying these currents, heat rise in connectors is not significant. Temperature rise can amount to several tens of degrees centigrade with circuits carrying ten or more amperes. In such cases, designers tend to specify alloys having conductivities in excess of approximately 40%IACS to minimize heat rise. To offset any increase in temperature, alloys having high

stress relaxation resistance in the 125-175°C operating range are also preferred to avoid loss in contact force.

Table 2

ELECTRICAL CONDUCTIVITIES OF SELECTED COPPER CONNECTOR ALLOYS

UNS Designation	0.2% Yield (ksi)	Min. Annealed Conductivity Electrical/Thermal (%IACS)/BTU/sq ft/ft/hr/°F
C151	55	95/215
C172	110	22/62
C1741	105	45/120
C510	100	15/40
C521	106	13/36
C7025	90	40/98
C724	110	10/28
C725	80	11/31

ALLOY FORMABILITY

Connector designs require that the alloy strip used be capable of being formed in bending, over radii which can approach a fraction of the strip thickness in dimension, and over angles of bending which range from a few degrees to nearly 180°. The limiting radius over which an alloy can be formed without fracture is usually dependent upon the relation of the axis of bending with the strip's rolling direction: longitudinal where the bending axis is normal to the latter, transverse where this axis parallels the rolling direction. The limiting radius of bending is usually greater for the transverse direction in copper alloys.

Limiting transverse bending radii for rolled temper and mill hardened alloys are illustrated by Figure 6. The precipitation hardened alloys are capable of providing the better strength-to-bend performance over rolled temper alloys, when higher strengths are needed. These alloys are also usually more isotropic in bend formability.

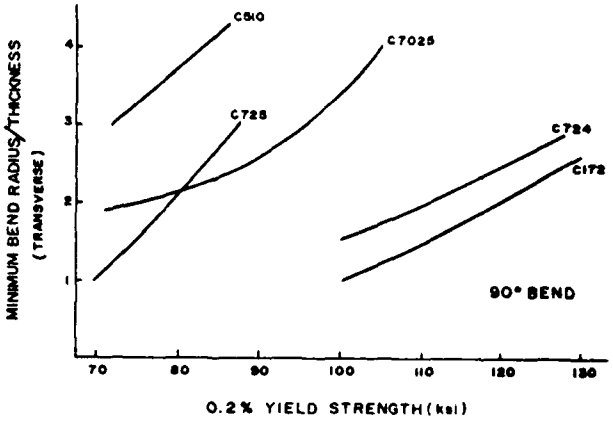


Figure 6: Strength-to-bend performance of selected copper alloys in the transverse orientation. Shown are test results for 90° bending in plane strain deformation.

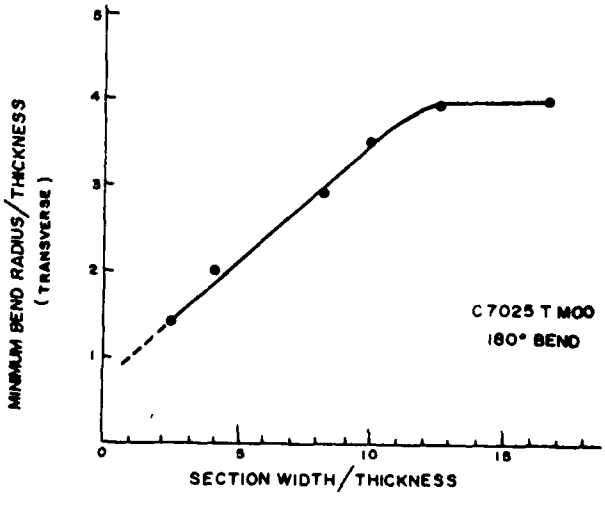


Figure 7: Bend formability improvement under plane stress deformation. Shown is the effect of decreasing the width: thickness ratio on the minimum bending radius over which strip can be formed without fracturing. The effect, shown here for 180° bending, holds also for forming over smaller angles.

The minimum bending limits depicted in Figure 6 were developed under conditions of plane strain (sample width/thickness ratio > 8). This data representation is typical for the industry and is useful in ranking of alloys. But, as every part designer and stamper knows, it is possible in practice to form strip to much smaller radii. Aside from subtle tooling effects, some portions of connectors are formed in plane stress where the width to be formed is only a few times the strip thickness. The improvement in bend forming limit as the width to be formed is decreased in dimension is illustrated by Figure 7. The extent to which part dimensions can improve bend formability should be determined for the specific alloy and temper of interest.

Bend formability is determined with tooling having discrete radii, and the limiting radii (normalized by dividing by thickness) are plotted as a smooth curve. Detection of the onset of failure when fine cracks must be discerned from deformation bands can add a measure of subjectivity to the measurement. The curves presented previously are based upon direct comparisons among alloys so as to minimize such subjectivity and thereby, permit relative ranking of alloys. Adding experience with one alloy to this ranking can be used to indicate which other alloys can be successfully formed into a particular part.

DESIGN CONSIDERATIONS

Initial contact force, as discussed earlier, is determined by the design geometry of the spring member and the load deflection response of the alloy being used. This contact force can decrease in time because of dislocation rearrangement. To predict the force remaining at the end of a connector's service life requires that the rate of stress relaxation for a particular alloy be known.

Alloys can be compared on terms of the actual applied stress that remains (at a prescribed service life) versus the minimum bending radius that is available for part fabrication (at that strength, or temper of the alloy). Earlier, stress relaxation was described in terms of a percentage of the initially imposed stress that remained after some test duration. The actual stress remaining and therefore, the force operating at the contact, can be calculated directly from percentage stress remaining data.

Stress relaxation performance (starting with an actual applied stress of 80% of the 0.2% yield strength) of several alloys, at their limiting bending radius (measured under plane strain condition for that yield strength) are compared in Figure 8. The precipitation hardened alloys

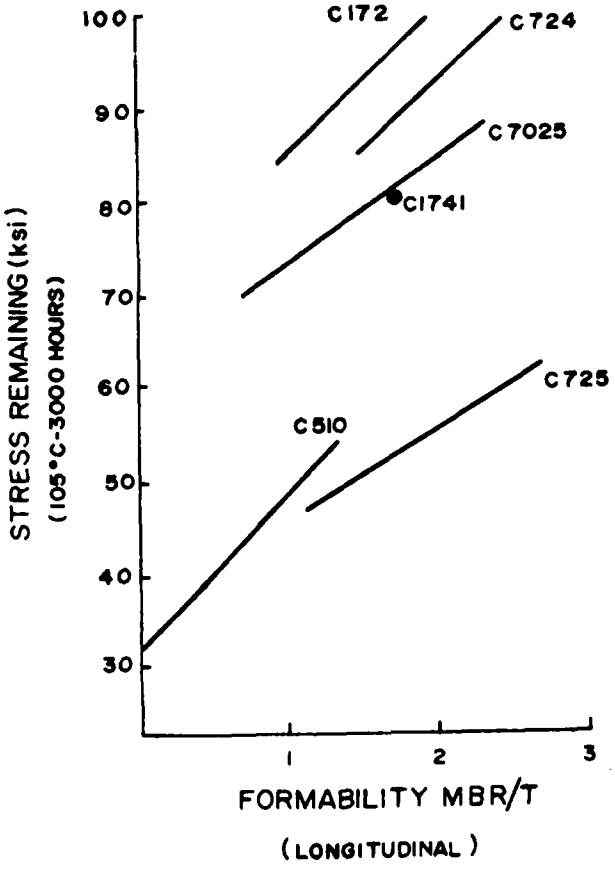


Figure 8: Stress relaxation resistance-bend formability comparison between several copper alloys in their longitudinal direction. The test samples were initially stressed to 80% of their respective yield strengths at each bending radius limit.

generally show the greatest stability, followed by stabilized tempers of alloys that are hardened by cold work, such as the phosphor bronzes.

Stress relaxation can, like other copper alloy strip properties, be dependent upon orientation relative to the rolling direction. The difference in stress relaxation resistance between longitudinal and transverse orientations, as measured by the percentage of the initial stress remaining, is typically around 5-10% (actual percentage) after 10 years at 105°C (221°F). Some alloys are more affected than others, and stabilization and precipitation heat treatments reduce this dependence. Stress relaxation resistance may be greater in either orientation, depending primarily upon the alloy.

Orientatation dependence of bend formability can be significant, especially for rolled-to-temper and stabilized alloys versus some precipitation hardened alloys which are more isotropic. Thus, the actual stress remaining shown in Figure 8 for the longitudinal orientation will be about the same in the transverse orientation for C172 and C724 (which have very nearly isotropic bend formability in the TM02 and TM04 mill hardened tempers). C510 and C725 can be more anisotropic, and these alloys have significantly poorer transverse bend formability in their stronger tempers. Accordingly, the transverse actual stress remaining at a given MBR/t will be less than shown in Figure 8 for the latter alloys, especially at the larger allowable bending limits associated with the stronger tempers.

Increasing temperature of exposure also results in decreasing actual stress remaining. The mill hardened alloys, with their stable dislocation structures, are more resistant to loss in applied stress than are the rolled temper alloys. The precipitation hardened alloys shown in Figure 8 (namely C172, C1741, C7025 and C724) continue to exhibit good stability at 150°C. In contrast, phosphor bronzes and C725, even if stabilized by low temperature heat treatment, do not provide adequate stability for use in applications above ~100°C.

CONCLUSIONS

Alloy selection is a vital part of connector design. When applied properly, the mechanical and physical properties presented in the preceeding paper can be used to predict connector functioning, and to also improve performance through designs which are more optimal for their defined use. The trend towards greater miniaturization, and more demanding operating conditions, places added emphasis on the need for alloy selection and the recognition of how their final processed condition impacts those properties which are crucial to connector performance.

ACKNOWLEDGEMENTS

The information contained on this review is representative of research efforts at Olin's Metals Research Laboratories over the several past years. Particular acknowledgement is given by the author to Eugene Shapiro and Jack Crane for their contributions to this area of technology and for discussions on the subject.

REFERENCES

1. P. Parikh, E. Shapiro, and J. Crane, Proceedings, CONNECTORS AND INTERCONNECTION TECHNOLOGY SYMPOSIUM, 416-434 (1976).

2. P. Parikh, S. P. Zarlingo, ibid, 435-449 (1976).

3. W. Brenneman, D. S. Larson, E. Shapiro, and R. Tatum, ibid, 281-298 (1979).

4. ASTM Standard Method E855-51

5. E. W. Filer and C. R. Scorey, STRESS RELAXATION TESTING, ASTM STP 676, 89-111 (1979).

6. A. Fox, Proceedings, CONNECTORS AND INTERCONNECTION SYMPOSIUM, 93-98 (1984).

7. J. F. Breedis to R. J. Hauser, Proceedings, ibid, 123-130 (1985).

8. H. Koeppen and L. Muller, ibid, 153-165 (1985).

9. J. F. Breedis and A. A. Khan, ibid, (1986).

10. E. Shapiro and J. Winter, RESIDUAL STRESS AND STRESS RELAXATION, 331-339, Plenum (1982)

LOW COST, HIGH RELIABILITY, SEPARABLE ELECTRONIC CONNECTORS

Robert P. Diehl
Burndy Corporation
Norwalk, Connecticut, USA

Abstract

Based on cost considerations tin or tin-alloy plated contacts are frequently being used in dry circuit electronic applications of low voltage and current. Traditionally high quality gold plating has been specified for these demanding applications because it does not form oxide or tarnish and therefore can provide clean metal-to-metal contact. The mere substitution of tin or tin-alloy plating for gold may result in unreliable contact interfaces. Inherent oxide or tarnish films present on tin or tin-alloy plating must be penetrated and dispersed during contact mating in order to establish clean metal-to-metal contact. The formation of metallic whisker outgrowths from pure tin plating on interconnections has to be addressed. After mating the tin or tin-alloy contact interface has to remain resistant to corrosive attack and be immune to the effect of "fretting corrosion."
This paper presents the results of research and testing experience which applies to the successful use of low cost, reliable, separable dry circuit tin-alloy plated contacts for electronic connectors.

WHEN CURRENT PASSES between two touching electrical contacts in an electronic connector it encounters an extra resistance due to the disruption of smooth current flow caused by the interface. This resistance is known as the contact resistance. The contact resistance has two parts, the constriction resistance and the film resistance. The constriction resistance is due to a purely geometric effect from the constriction of the lines of current flow through the tiny areas of true contact. The film resistance arises from reduced or no current flow through any surface films which may be present. Surface films can be the result of corrosion or organic in nature resulting from polymerization reactions.

In order to achieve an initially low contact resistance sufficient contact area has to be developed at a clean metal interface. If the contacts are film covered, the films have to be penetrated first and then sufficient metal contact established.

Surface film breakthrough can occur either mechanically, during the mating of contacts, or electrically. If the electric field strength is insufficient to penetrate any film in the contact interface then a dry circuit exists. The voltage level upon which a dry circuit exists for most films is defined as one with an open circuit voltage of 20 millivolts or less. This paper will deal only with contact interfaces under dry circuit conditions.

Dry circuit electrical contacts then must function in one of two ways. Either films must be absent initially and must not be formed during the lifetime of the contact. High quality gold plated contacts function this way. On base metal plated contacts films may be present initially but they must be penetrated mechanically by the contact mating process[1], and the contact interface must then be impervious to ingress by corrosive gases in the atmosphere, be gas-tight.

The contact interface also must be immune to the effects of fretting corrosion.[2-4] Fretting corrosion results from repeated rubbing of mated contacts over distances of approximately 1.2 to 125 μm (0.05 to 5 mil). This may result from vibration or from movement during temperature changes due to the difference in thermal expansion coefficients of the connector materials and supporting structures. The wear debris from fretting oxidizes and remains at the contact interface, eventually producing large increases in contact resistance.

To develop a low cost, high reliability, base metal separable electronic connector system the following considerations have to be implemented. Contact plating selection and design of the contact and connector have to assure the penetration of resistive surface films and provide a large gas tight contact area which is immune to the effects of fretting corrosion.

SELECTION OF THE CONTACT PLATING

Traditionally the outer surface plating of electrical contacts for electronic connector applications has been gold plate. Gold has many desirable properties as a contact material, most notably its complete freedom from oxide films. Prior to gold, silver was the preferred contact metal because of its conductivity. The first use of gold was for the purpose of protecting the silver from tarnish, to improve its solderability. To do this effectively it soon became clear that fairly thick and expensive platings were needed and that the silver plate under the gold was not really performing a useful function. In the early 1960's the main thrust of research in contact materials was in optimizing underplate systems other than silver[5], reducing gold plate porosity[6], and the application of contact inhibitor - lubricants.[7] As a result of these developments, gold plated contacts with reasonable thickness of plate perform satisfactorily in a wide variety of applications. Nevertheless, the escalating cost of gold in the 1970's continuing till now, even with thin and selective locating of platings in the mating area of a contact, provides sufficient economic incentive to consider its total elimination with a base metal replacement.

Basic research on the physics of surface contact led to the conclusion that the hardness of a surface layer relative to the bulk material affects the degree of true surface contact.[8] Deposits harder than the substrate result in less true contact area. Soft plate gives almost complete true contact[9], resulting in a gas tight, corrosion - resistant interface. Further, a soft base metal plate with a thin, brittle surface corrosion film which can be easily penetrated mechanically is desired. Soft, pure tin develops thin, brittle self-limiting corrosion films and its theoretical constriction resistance ranks close to gold.[10-11] Soft, tin-lead plate also forms thin brittle self-limiting corrosion films. A field exposure study in 1964 by the ASTM B-4 Committee reported on mated and unmated crossed rod contacts of various materials. They rated tin-lead plate the most corrosion resistant non-noble metal plating of those tested and place it very near gold.[12] We also tested tin-lead plating using the automatic contact resistance probe during our investigations of suitable reagents for gas-tightness testing which will be discussed later. Figure 1 gives contact resistance vs. load data for tin-lead plated coupons exposed unmated to various laboratory atmospheres. Except for fumes from concentrated nitric acid (used for destructive gas-tightness testing only) the atmospheres tested produce films which were readily penetrated at loads of 150 grams or greater.

In the selection of a base metal contact surface plating one important consideration has to be addressed, whisker growth. Pure tin plating on interconnections can spontaneously grow, at some indefinite time, thin metal filaments (whiskers) from its surface.[13] The fragile hairlike filaments are single-crystal pure metallic outgrowths.

The metallic whiskers can grow to sufficient length to short circuit adjacent low-voltage level circuit paths (Figure 2). Also, breakage and relocation of the whiskers can cause electrical malfunctions by bridging circuit elements. The elimination of potential whisker growth is essential for electronic components. Whiskers are commonly straight or kinked and grow outward from their base. Figure 3 is a scanning electron micrograph of typical whisker growth on contacts.

Due to the unpredictable nature of spontaneous whisker growth the time required to initiate natural growth may vary from weeks to months. The application of external pressure is known to vastly accelerate the incubation and growth process. Pressure therefore, is a means to quickly test various tin and tin-alloy platings in order to determine their whiskering potential. A transmission electron shadowgraph of whiskers from an accelerated pressure tested matte tin deposit is shown in Figure 4. Figure 5 is a transmission electron shadowgraph of an accelerated pressure tested tin-lead plate. Complete whisker supression has been achieved. The observation of additional accelerated pressure tested specimens and specimens allowed time to spontaneously grow whiskers has shown that tin alloyed with lead is an effective method of eliminating whisker growth. As little as 2 weight percent lead alloyed with tin virtually eliminates spontaneous whisker growth.

The tin-alloys we have selected for our low cost, high reliability, separable electronic connectors are ones alloyed with 2 to 40 weight percent lead depending on the application. For an application where the contacts have solder tails that are terminated to a printed circuit board 60/40 Sn-Pb would generally be considered. Some special applications may require greater wear resistance or a higher softening temperature to prevent unwanted plating reflow. For these cases the lead content may be reduced to as little as 2 weight percent.

CONTACT DESIGN

The separable base metal connection concept* consists of a wedge-shaped contact plated with 5 to 10 μm (0.2 to 0.4 mil) tin-lead plate and a mating normal force of a minimum of 150 grams.

Figure 6 is a photomicrograph of a micro-section through the contact interface of a wedge contact mated to a printed circuit board pad. As we can see in this cross-section the degree of deformation caused by the mating wedge contact is sufficient to break-up surface films and contamination and achieve true metal-to-metal contact. The tin-lead contact plating appears as white tin-rich areas and black lead-rich areas as a result of chemical etching for structural detail enhancement. The ductile nature of the tin-lead surface plating results in extensive extrusion and redistribution of metal within the contact area, especially toward the periphery of the joint, thus significantly increasing the area of contact. As a result the wedge contact maintains a metallurgically sealed electrical interface resistant to the ingress of corrosive contaminants across the contact area. The high compliance of the tin-lead plate allows maintaining the contact area as the load on the contact members decreases.

For example, increasing and decreasing loads in a given mating cycle can be characterized by "breathing" and "unmating" modes as shown in Figure 7. In the breathing mode loading and unloading occurs without contact separation between mating cycles. In the unmating mode the contacts are separated and the contact position changed prior to reloading. The bar graphs are for contact resistance vs. load for unplated and tin-lead plated wedge contacts. Unplated contacts show a considerable degree of elastic recovery resulting in a decrease of contact area and non-compliance in the decreasing load half of the cycle for both breathing and unmating modes, and this is repeated during subsequent cycles. This is shown by the increase in contact resistance as the load is decreased. Wedge contacts with the ductile compliant tin-lead plate are completely in the plastic deformation regime. All the contact area which is generated at the maximum load in the first cycle is preserved in subsequent breathing cycles as long as the contacts are not separated. If they are separated and remated, the first half of the new mating will again generate a maximum contact area which remains constant during subsequent load fluctuations. This is shown by the little or no change in contact resistance as the load is decreased. The inherent nature of the tin-lead plate allows the formation of large areas of true metal-to-metal contact initially and then maintains the contact area with decreasing contact force thus establishing a highly reliable base metal contact interface.

* GTH™, **BURNDY** Corp., Norwalk, CT. 06856. **GTH** is an acronym for Gas Tight High Force.

GAS TIGHTNESS TESTING - Gas tightness is a physical property of the contact interface in a mated contact pair. A gas-tight interface excludes ingress of air pollutants. It follows that a gas-tight contact pair will continue to conduct even if exposed to pollutants that produce a totally non-conducting corrosion film on an unprotected metal surface. A convenient gas-tightness test for tin and tin-alloys is exposure of a mated contact pair to fumes of concentrated nitric acid for 1 hour followed by exposure to fumes of ammonium sulfide for 15 minutes.[14] Contact resistance is measured before and after exposure. Figure 8 illustrates typical results of such a test. The unmated and lightly loaded contacts failed the test, thus demonstrating the ability of this procedure to detect a non-gas-tight connection. However, contacts mated at 150 grams or more produce highly reliable gas-tight interfaces. Figure 8 shows that when contacts after 10 mating cycles are exposed mated they are able to maintain a fully gas-tight connection.

FRETTING CORROSION TESTING - Small amplitude rubbing of non-noble metals in contact may result in an accumulation of oxide film in the contact area due to acceleration of oxidation by frictional heat. This can give rise to high contact resistance values. All non-noble metals exhibit this effect, which is prevented by contact lubrication.[3,15-16]

We have investigated the effect of contact geometry on the susceptibility of base metal plated contacts to fretting corrosion. In these experiments wedge contacts were compared with rounded or conventional contacts. Gold plated contacts used for comparison were always of conventional geometry as is the case in practical connector design. These experiments have already been reported in detail.[17] Figure 9 summarizes the result of an experiment that compared the wedge geometry with tin-lead plate to hemispherical contacts with either tin-lead or gold plate. In this experiment, relative motion was forced at the contact interface in order to isolate the effects of geometry. Although not producing total immunity with forced movement wedge-shaped contact geometry has significant fretting corrosion resistance without lubrication. Time to failure is similar to that for 0.75μm (30μ in.) of hard gold on conventional contacts. Failure of gold plated contacts is due to wear through of the gold. Fretting corrosion then occurs with the underlying base metal beneath the gold. In actual connectors the combination of correct geometry and proper structural support for the contacts can provide total immunity to fretting corrosion without the necessity for contact lubrication.

Any base metal contact plate can be made to fail under forced fretting testing. Most of the literature which has been published to date on the subject of fretting corrosion have been reports of experimental research on model contacts subjected to forced motion. Up to this date, there has been no fretting corrosion test on actual connector assemblies which has been accepted by any authoritative industry committee. Fretting tests should simulate and accelerate the actual use conditions of the connectors. Our position on this subject is as follows: (1) Vibration testing of complete mated connector assemblies, mounted as they would be in actual use, is a reliable and convenient test procedure for fretting corrosion susceptibility. Equipment utilized is the same as that used for MIL spec. vibration testing. (2) Temperature cycling testing of complete mated connector assemblies may also induce fretting corrosion by imparting small periodic interfacial motion due to differential thermal expansion. The test conditions for vibration and temperature cycling testing should simulate and accelerate the actual vibration levels and temperature excursions of the equipment in which the connectors are used.

In order to detect fretting corrosion damage by a vibration or thermal cycling test, it is necessary to perform dry circuit contact resistance measurements before and after the test. An irreversible rise in static contact resistance indicates that fretting damage has occurred. The monitoring of dynamic contact resistance to measure transient resistance changes under fretting conditions has also been reported.[4,18-19]

The diagnosis of fretting corrosion is confirmed by microscopic examination of the contact interface after unmating. Fretting corrosion appears as grey or black areas in the contact interface. Figure 10 and 11

illustrate the typical appearance of forced motion fretting corrosion[17] on flat coupons rubbed with tin-lead plated hemispherical and wedge contacts respectively. The penetrating wedge geometry acts to confine the oxide film to the periphery of the contact area and to keep enough of the contact clean so that low resistance values are maintained.

FIELD EXPERIENCE

Serious failures in the field have been reported for IC Sockets with tin-alloy-plated contacts, especially when mated to gold plated leads.[20-21] The contacts were similar in design to conventional gold-plated socket contacts except that tin-alloy was substituted for the gold. Contact forces were variable and some were much lower than those used for wedge-geometry contacts or generally recommended for tin or tin-alloy-plated contacts.[22]

On the other hand, IC Socket connectors with tin-lead plate and wedge contacts have been in use since 1972, and billions have been in service for many years. No corrosion failures of any type have been reported. This applies to connectors mated to gold as well as to tin alloy.

Table 1 presents the results of a two-year program during which connectors and contacts were exposed to field sites with known pollution levels. These sites are maintained and monitored by Battelle Laboratories, Columbus, Ohio. The samples were prepared and submitted by BURNDY and are representative of normal production runs. The various contact and connector (Con.) categories are described in the footnotes of Table 1.

Categories Au/Au and Au/Au (Con.) contrast the effects on unshielded contacts and on those protected by the connector body. Both categories were worse in Site B. A possible explanation is that the higher chloride levels in Site B react more severely with the nickel underplate on these contacts. Categories Sn/Sn-Pb, Au/Sn-Pb and Sn-Pb/Sn exhibited the same stability at the monitored field sites as that experienced with the wedge contact system in actual use.

The use of like-to-like plating on contacts is recommended whenever possible. However, when this is not feasible (for example, when IC Sockets with tin-alloy-plated contacts are used with gold-plated IC leads) wedge-geometry, high-force contacts have given satisfactory service.

The successful use of the wedge contact system has been reported in an actual product application.[23] The minimum normal force of 150 grams required for reliable performance of tin alloy interfaces has also been recommended.[19]

SUMMARY AND CONCLUSIONS

The low cost, high reliability, separable base metal electronic connector concept is the culmination of a unique approach to develop a contact system with equivalent reliability to high quality gold contact systems. The simple substitution of base metal contact plate for non-gas-tight interfaces has not been pursued. The primary elements of the concept which have been developed to assure contact reliability are: (1) a defined wedge-shape contact mated at a high normal force to penetrate resistive films and stabilize the contact interface from movement thus facilitating fretting corrosion resistance and (2) the contact plating material. A 5 to 10 μm (0.2 to 0.4 *mil*) thick ductile compliant tin-lead surface layer provides a "gas-tight" contact interface when mated at normal forces of 150 grams or greater. The tin-lead plate has intrinsic corrosion resistance and is immune from whiskering.

Field experience of this contact system has demonstrated its high reliability.

REFERENCES

1. Osias, J.R. and J.H. Tripp, Wear. 9, 388-397 (1966)
2. Bock, E.M. and J.H. Whitley, Proc. Holm Seminar on Electric Contacts. 20, 128-138 (1974)
3. Antler, M., Proc. Annual Connectors and Interconnection Technology Symp. 16, 232-242 (1983)
4. Abbott, W.H. and K.L. Schreiber, Proc. Holm Seminar on Electrical Contacts. 27, 211-219 (1981)
5. Garte, S.M., Plating. 55, 946-951 (1968)
6. Garte, S.M., Plating. 53, 1335-1339 (1966)

7. Antler, M., S.M. Garte and S.J. Krumbein, Proc. Eng. Seminar on Electric Contact Phenomena. 13, 79-92 (1967)
8. Pullen, J. and J.B.P. Williamson, Proc. Royal Soc. A. 327, 159-173 (1972)
9. Tripp, J.H. and S.M. Garte, Proc. Holm Conference on Electrical Contacts. 26, 227-236 (1980)
10. Germer, L.H. and J.L. Smith, Bell Systems Tech. J. 36, 769 (1957). See also Germer, L.H., J. Appl. Phys. 25, 332 (1954)
11. Mc Knight, L.G. and J.M. Sawina, Phys. Rev. A2, 762 (1970)
12. Blake, B.E. "Summary Report of the ASTM Section G, Contact Field Tests", ASTM Section G, Committee B-4, Sub-Committee IV, 1964.
13. Diehl, R.P. and N.A. Cifaldi, Proc. Annual Connector Symp. 8, 328-336 (1975)
14. Garte, S.M. and R.P. Diehl, Proc. Annual Connector Symp. 9, 239-249 (1976)
15. Freitag, W.O., Proc. Holm Seminar on Electrical Contacts. 21, 17-23 (1975)
16. Freitag, W.O., Proc. Holm Seminar on Electrical Contacts. 22, 57-63 (1976)
17. Garte, S.M., Proc. Holm Seminar on Electrical Contacts. 22, 65-70 (1976)
18. Abbott, W.H., Proc. Holm Conference on Electrical Contacts. 29, 55-59 (1983)
19. Currence, R. and W. Rhoades, Proc. Holm Conference on Electrical Contacts. 29, 81-103 (1983)
20. Mills, G.M., IEEE Trans., Components, Hybrids, and Mfg. Tech., CHMT-2, No. 4, 476-489 (1979)
21. Mottine, J.J. and B.T. Reagor, Proc. Holm Conference on Electrical Contacts. 29, 61-71 (1983)
22. Abbott, W.H., M. Antler, J.H. Whitley, G. Mills, J. Long, M. Peel, and G. Saxenmeyer, Proc. Holm Conference on Electrical Contacts. 27, 287-322 (1981)
23. Tuck, J., Circuits Manufacturing. July, 69-70 (1984)

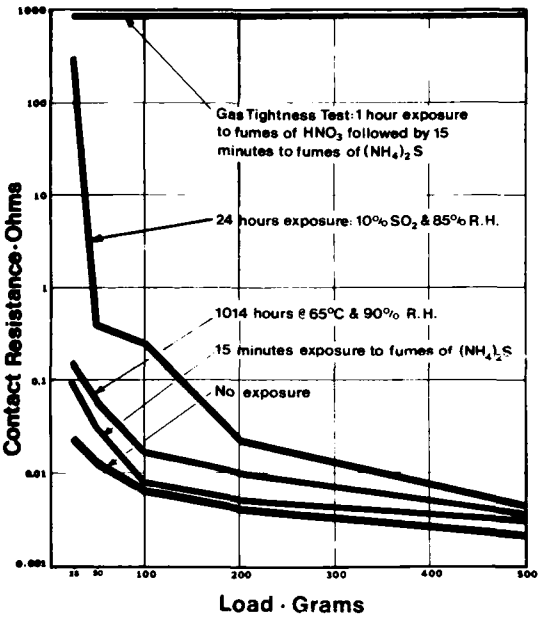


Fig. 1. Contact resistance vs. load of tin-lead plated coupons exposed unmated to various laboratory atmospheres.

Fig. 2. A scanning electron micrograph of a single whisker bridging two contact surfaces. 70X.

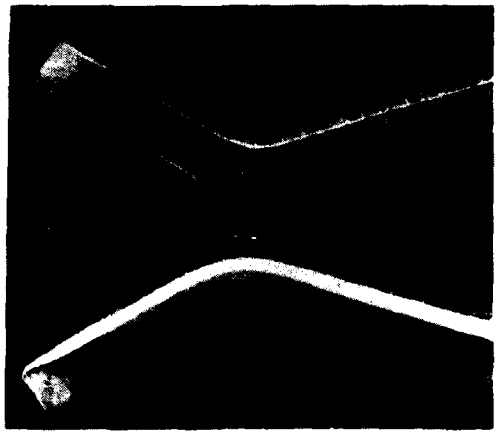


Fig. 3. A scanning electron micrograph of typical whisker growth on contacts. 35X.

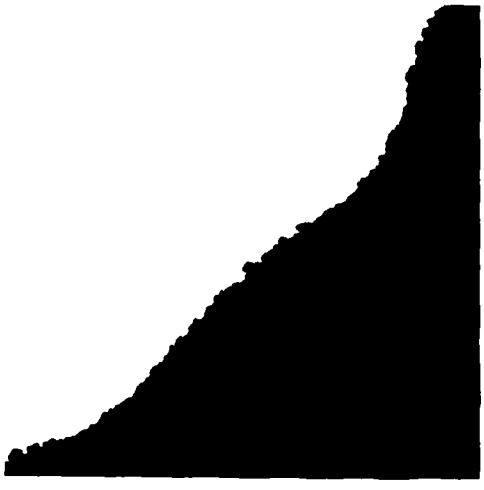


Fig. 5. A shadowgraph of an accelerated pressure tested tin-lead plate which showed no whiskers. Complete whisker supression has been achieved. 4250X.

Fig. 4. A shadowgraph of whiskers from an accelerated pressure tested matte tin deposit. 850X.

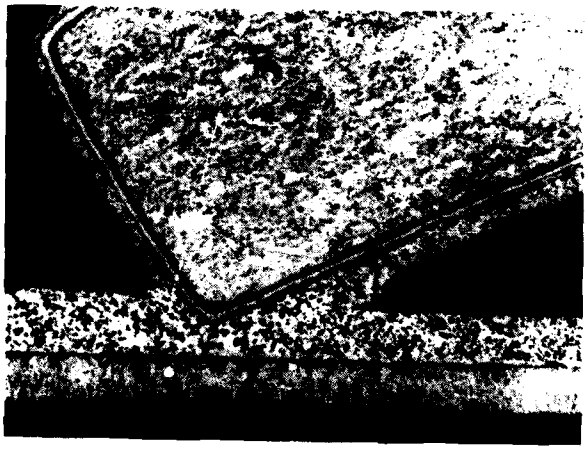


Fig. 6. Micro-section through the interface of a wedge contact with 5 µm (200 µin) of 60 percent Sn/40 percent Pb and 150-g force. 210X.

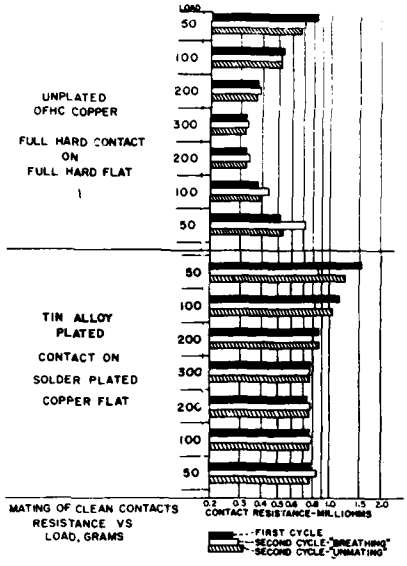


Fig. 7. Contact-resistance vs. load characteristics of unplated and tin-lead plated wedge-shaped contacts.

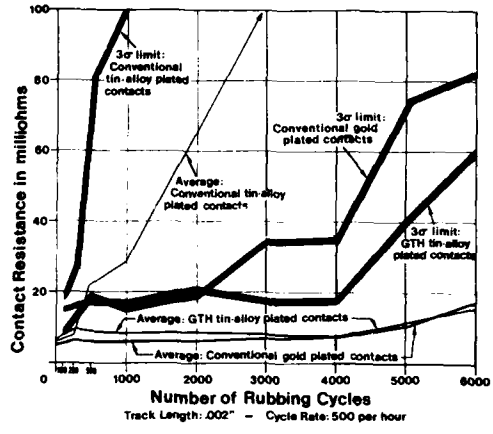


Fig. 9. Contact resistance vs. number of rubbing cycles along a 50 μm (2 mil) track at 150g force and 500 cycles/hr. The light lines are the average values. The heavy lines are the 3σ limits (3σ x standard deviation; sample size of 25). Any individual value will not exceed this limit with a statistical confidence level of 99.8 percent.

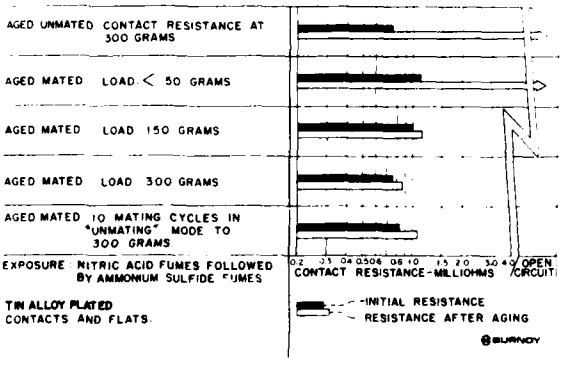


Fig. 8. Gas-tightness test data for tin-alloy-plated contacts and flats before (black bars) and after (white bars) exposure to fumes of nitric acid (1 hr.) and ammonium sulfide (15 min.).

Fig. 10. Fretting corrosion failure on a flat rubbed with a hemispherical contact for 200 cycles. 60X.

Fig. 11. Fretting corrosion failure on a flat rubbed with a wedge contact for 6000 cycles. Arrows show length and direction of forced motion. 25X.

Table 1
Contact Resistance (Milliohms) After Two-Year Exposure

Site	Sample Category	Initial Average	Final Min.	Final Max.	Average	Sample Size	Standard Deviation
A[a]	Au/Au[c]	1.6	1.1	2.4	1.5	28	0.26
A[a]	Au/Au (Con.)[d]	4.7	3.9	5.5	4.8	35	0.45
A[a]	Sn/Sn-Pb[e]	4.7	4.3	5.9	5.3	14	0.39
A[a]	Au/Sn-Pb[f]	11.0	6.6	11.6	9.8	16	1.31
A[a]	Sn-Pb/Sn[g]	6.7	6.1	7.4	6.7	16	0.37
B[b]	Au/Au[c]	1.6	0.9	999	75.5	27	200.4
B[b]	Au/Au (Con.)[d]	4.8	3.9	10.6	5.5	36	1.35
B[b]	Sn/Sn-Pb[e]	5.0	5.0	6.4	5.9	14	0.4
B[b]	Au/Sn-Pb[f]	11.0	11.2	14.7	12.3	16	0.87
B[b]	Sn-Pb/Sn[g]	6.8	7.8	9.8	8.4	16	0.58

[a]Site A is an air traffic control facility near Pittsburgh, PA. It has considerable chloride pollution and temperature-humidity variations.
[b]Site B is a paper mill in Maine with high sulfide and chloride pollution and temperature-humidity variations. It is one of the most severe sites in the field-test program.
[c]Stamped and formed pin-and-socket contacts with gold plate 1 μm (40 μin.) in thickness. These contacts are mated but not installed in connector bodies.
[d]Same as contacts described in Footnote c except that these are installed in connectors and mated.
[e]IC package with tin-plated leads plugged into dual in-line IC socket with tin-alloy-plated contacts.
[f]Same as contacts described in Footnote e except that the IC package has gold-plated leads.
[g]Tin-alloy-plated contacts in a connector mated to a flexible circuit with tin-plated pads.

THE PERFORMANCE OF ELECTRONIC CONNECTORS IN FLOWING MIXED GAS LABORATORY ENVIRONMENTS

W. H. Abbott
Battelle Columbus Division
Columbus, Ohio, USA

ABSTRACT

Studies were made on the statistics and kinetics of degradation of contact resistance on several generic types of commercial connectors exposed to flowing mixed gas laboratory environments. Most studies were made in what has been termed the Battelle Class II FMG environment (H_2S-NO_2-Cl_2). Comparative data were obtained on several of the commonly-used H_2S, SO_2, and H_2S-SO_2-NO_2 environments.

Results have clearly shown that statistically significant degradation may occur in laboratory environments containing only part per billion (ppb) levels of pollutants. However, this requires the incorporation of the "correct" sulfide-chloride interactions as found in field environments.

Two additional points were demonstrated in these studies. First, the addition of SO_2 in the H_2S-NO_2-Cl_2 environments produces no statistically significant change in reaction kinetics or mechanisms. Second, comparative data for exclusive sulfide environments typically produced extremely low acceleration and unrealistic reactions in comparison to field experience.

EARLIER PAPERS BY THE AUTHOR [1,2] have introduced the concept of classes of field environments relative to actual electronics operating environments (indoors and/or inside enclosures). These studies demonstrated that environments could be classified on the basis of (1) reaction kinetics and chemistries on standard materials such as copper, and (2) reaction mechanisms and chemistries on materials such as porous gold platings. The first of these concepts has already evolved as a U.S. national standard [1].

This work, together with other relevant studies [4-9]' have clearly defined the importance of chlorides, and more specifically, the sulfide-chloride interactions, in controlling indoor corrosion processes. Other work, as summarized in reference [1] has shown that these interactions exist with varying but high frequencies of occurrence in most types of operating environments worldwide.

In recent years it has become clear that laboratory test environments for use in reliability studies must adhere to at least two principles. First, they must be tailored to the probable range of application environments. In other words, a single test cannot suffice for the simulation of all conditions. Second, realistic and useful tests must properly incorporate the interactions as found in field environments.

These considerations have led the author to propose three test environments for accelerated test purposes. These have come to be known as the Class II, III, and IV FMG environments. By definition, Class I test does not exist since field studies have shown precedent for corrosion-related degradation in this class of environments. Details concerning these tests, methods of control of the tests, and expected results are given in reference [1].

Most published work concerning these or similar tests have dealt with the reactions on metallic surfaces in the form of coupons. While such data are extremely useful for calibration and verification of tests, they provide little insight into the likely effects of such test environments on hardware performance. A discussion of this subject is timely since the FMG procedures are being widely adopted, at least in the United States, for reliability studies. The purpose of this paper is to present some of the first data on the application of FMG procedures to connector reliability studies. It was specifically

designed to examine degradation rates on several generic types of connectors and with sample sizes sufficiently large to develop statistically meaningful results. The studies to be discussed in this paper were restricted to the relatively mild Class II FMG environment.

EXPERIMENTAL PROCEDURES

MATERIALS - Three types of connectors were used in this study. Two may be described as edge-card connectors. Both represented conventional designs with relatively low degrees of "shielding". They differed mainly in the contact metallurgy on the connector springs. One was a conventional, cobalt-hardened, acid gold over sulfamate nickel. The second had a gold-flashed, palladium-nickel metallurgy.

The third connector was a conventional pin and socket connector. In the mated condition, this configuration presented a high degree of shielding from the environment. However, it also had a high porosity plating in the contact area on the sockets.

A complete characterization of these connectors is given in Table 1. This includes actual measurements of plating thicknesses in the contact regions, porosity levels on all plated finishes, and the contact normal force. These results show that porosity was present to various degrees in all systems. These values represented real data obtained on production hardware.

Environmental Test System - The test system, operating conditions, and control methods were identical to those described in reference [1]. Due to the large numbers of samples used in this study, the chamber which was actually used had an internal volume of 3,600 liters.

Operating conditions were controlled at nominal Class II levels. These were:

- 10 ppb H_2S
- 10 ppb Cl_2
- 100 ppb NO_2
- 70 percent RH
- 30°C

These values represent target levels as actually analyzed within the test environment. Verification was obtained by the use of copper and porous gold reactivity coupons. Actual results obtained in all runs were within the expected response range, as shown in reference [1], Figures 5 and 6.

TABLE 1. CHARACTERISTICS OF CONNECTOR HARDWARE STUDIED

Edge Card #1

Connector
- Materials 1.4μmAu/0.88μmNi/Cu-Sn
- Porosity 10-15% of Springs >0 Pores
  0-2 Pores per Spring
- Normal Force - 212 ± 20 grams

PC Board
- Materials 1.6μmAu/8.0μmNi/Cu
- Porosity - approx. 90% of Tabs >0 Pores
  0-8 Pores per Tab

Edge Card #2

Connector
- Materials 0.1Au/1.9Pd-Ni/Cu-Sn
- Porosity - approx. 0
- Normal Force - 126 ± 15 grams

PC Board
- Materials 1.9μmAu/3.1μmNi/Cu
- Porosity - approx. 100% of Tabs >0 Pores
  5-11 Pores per Tab

Pin and Socket

Pin
- Materials 1.0μmAu/1.6μmNi/Cu-Sn
- Porosity - 100% of Pins >0 Pores
  2-6 Pores per Pin

Sockets
- Materials 0.9μmAu[a]/0.25μmNi[a]/Phosphor Bronze
- Porosity - 100% of Sockets >0 Pores
- Normal Force - 110-120 grams

(a) Values at Nominal Contact Interface

Measurement Procedures - Contact resistance change was used as the method of evaluating performance. All samples were measured initially and then after various intervals of exposure. For this purpose, samples were physically removed from the test environment and measured on high-speed, scanning and data acquisition systems. Typical pin counts actually measured at each interval were between 1000 and 3000.

Measurement conditions were 50 millivolts open-circuit voltage, 1 milliamperes D.C. current. At each measurement point, two measurements were actually made. One will be termed an undisturbed measurement taken directly

after removal of samples from the test chamber. Thereafter, each sample was mechanically "disturbed" to cause a slight degree of relative motion at the contact interface. This was done manually to produce a single motion event. The data actually presented at each interval will represent the composite of the undisturbed and disturbed readings.

Most of the studies were made on mated connector systems. However, some comparative data were obtained with one half of each of an independent sample group exposed in the unmated condition. In these cases, each sample was mated once for measurement purposes.

EXPERIMENTAL RESULTS

Coupon/Control Reaction - Figures 1 and 2 show, together with other data, the expected (and actual) results obtained on coupon samples freely exposed as controls in the test environments. Copper samples were analyzed by cathodic reduction and the porous gold samples were analyzed by contact resistance probing.

Figure 2 also shows data for several other types of test environments. These data will be particularly useful for comparing connector response in similar environments. Figure 2 also provides a concise summary of the accelerating effects of sulfide-chloride interactions. It further shows that typical, single gas exposures, even at relatively high concentrations, may be quite benign.

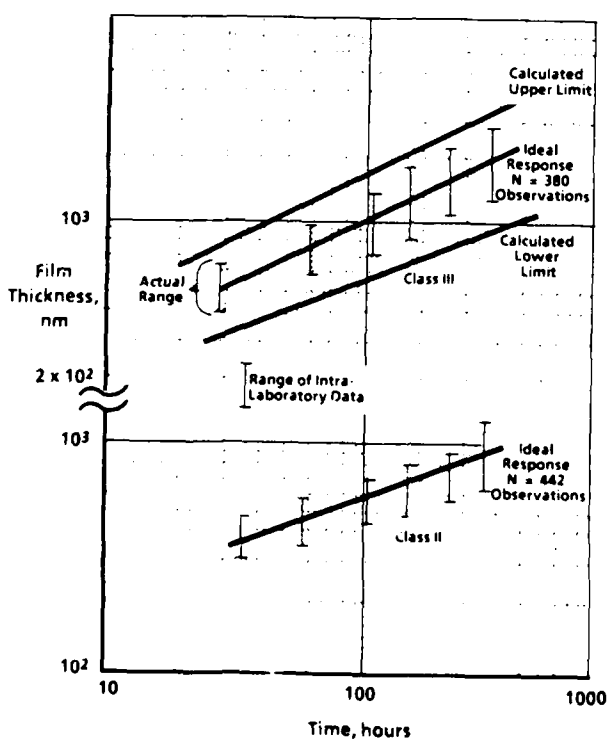


Figure 1. Corrosion Response Range of Copper Controls in FMG Mixtures

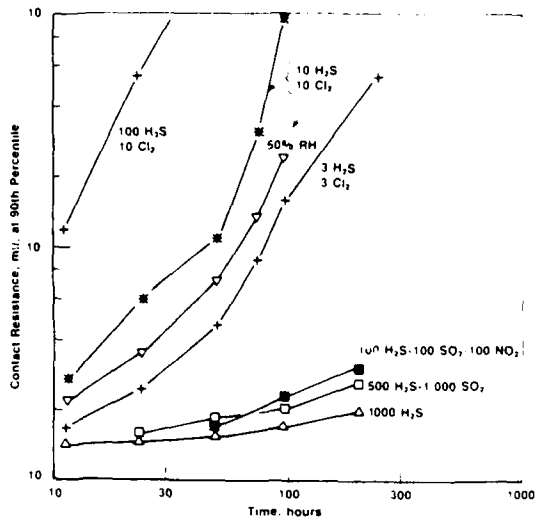


Figure 2. Kinetics of Porous Gold Degradation in Several Laboratory Environments; 30°C. 70% RH

CONNECTOR PERFORMANCE

Edge Card Connectors - Figure 3 shows the results for the two edge card connector systems for three exposure times. Two characteristics are shown in these data which are typical of virtually all mated connector systems. First, at any point of observation, the percentage of contacts which have actually degraded to a measurable or significant degree is typically small. For the purposes of this paper, the term measurable refers to the minimum level of ΔR, which can be reported in a practical test system. This is probably in the range of 2 to 3 milliohms. The term significant refers to levels taken as failure thresholds for test purposes. This may typically be in the range of 10 to 20 milliohms.

With respect to the latter limits, the percentage of degraded contacts is typically less than 1 percent and often less than .01 percent within meaningful test periods. These levels are, however, highly significant with respect to reliability implications, as will be discussed in a later section. For the moment, however, these results lead to the conclusion that for testing of this type, large sample sizes must be used before degradation can typically be detected.

The results in Figure 3 also address the kinetics of degradation. This is typically slow, as shown by the time to reach measurable levels of degradation. This leads to the conclusion that tests of this type cannot be conducted quickly. This is consistent with presently understood concepts of acceleration factors for such tests.

The results in Figure 3 show a dramatic attenuation in comparison to the results for coupons shown in Figure 2, even though both contact metallurgies are nearly the same. This

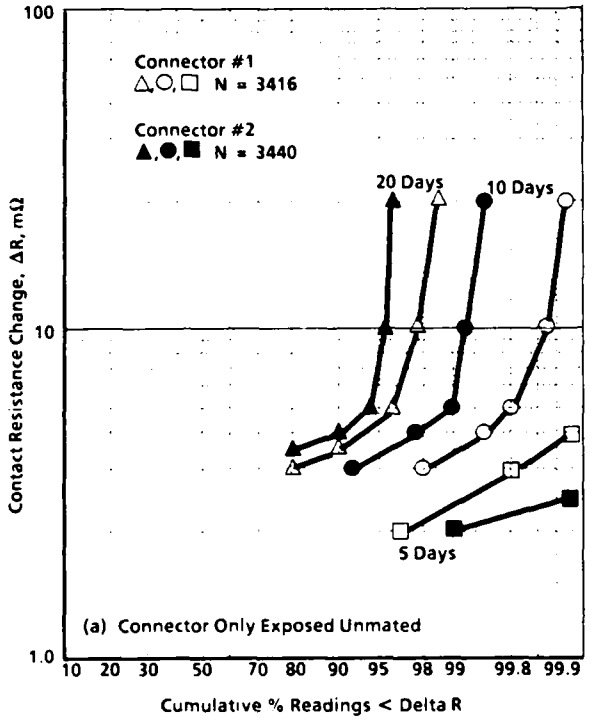


Figure 3. Performance of Mated Edge Connector(a) in Class II FMG Environment

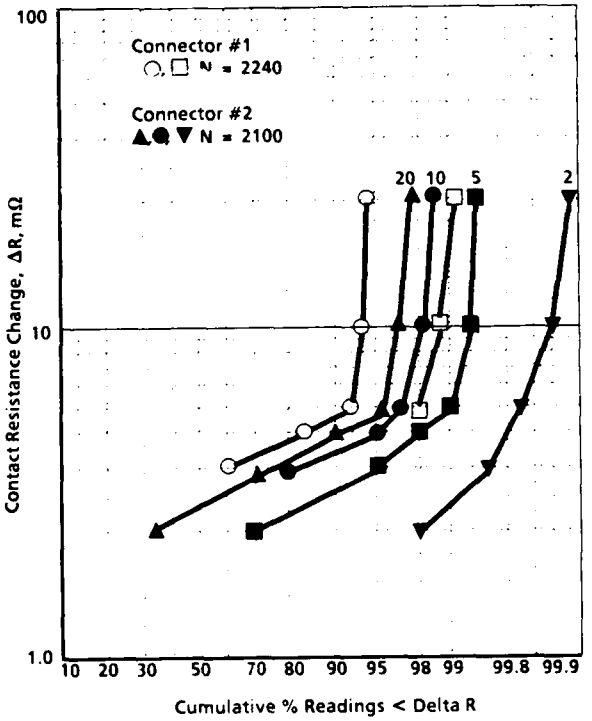


Figure 4. Performance of Unmated Edge Connectors in Class II FMG Environment

is largely a reflection of the well-known shielding effect, which is quite dramatic even for the open edge card connector construction.

Finally, Figure 3 compares the relative performance of the two contact metallurgies in connectors 1 and 2. This shows a considerably higher, long-term rate of degradation for gold-flashed palladium-nickel in comparison to gold over nickel. Although several factors may explain this behavior, the important features, for purposes of this paper, can be found in the comparative test data from exclusive sulfide environments.

Figure 4 shows comparative data for unmated exposures of the edge card connectors. In these experiments only the connector portion of the system was exposed to the test environment.

These data show far higher rates of degradation in comparison to mated connectors, although still far less than rates found on unshielded coupons. It is believed that the lower rates for the unmated connectors can be attributed to two factors. One is some degree of shielding by the plastic housing over the connector body. The second is contact design, together with contact wipe.

One final and important conclusion concerns the implication of these data to unmated connector performance in the field. They indicate that field life may be relatively short (1 to 3 years) in operating environments at Class II levels or above.

Pin and Socket Connectors. Figures 5 and 6 show comparable performance data for the pin and socket connector system. The mated data show better performance than was shown for the edge card system, in spite of a lower contact normal force and higher porosity levels (socket contacts). This, again, illustrates the important effect of environmental attenuation or shielding by the connector housing.

This effect is further illustrated in Figure 6, which shows unmated data obtained in experiments in which either pins or sockets were independently exposed to the test environment. Considerably higher degradation rates were observed on pin exposures (less shielding). This was found in spite of the lower porosity levels on the pins relative to the sockets.

Both data sets, however, show very high, long-term degradation.

Effects of SO_2. This important outdoor pollutant has normally not been included in the FMG environments. The reasons for this are discussed in reference [1] but, in summary, that work showed no added or statistically significant effect of SO_2 on coupons. Similar studies have now been made with connector hardware. Sulfur dioxide was added as a third constituent to the Class II mixture. The concentrations used in independent experiments were 100 and 300 ppb. It should be noted that these values are proportionately far higher than the concentrations of H_2S or Cl_2. They are also

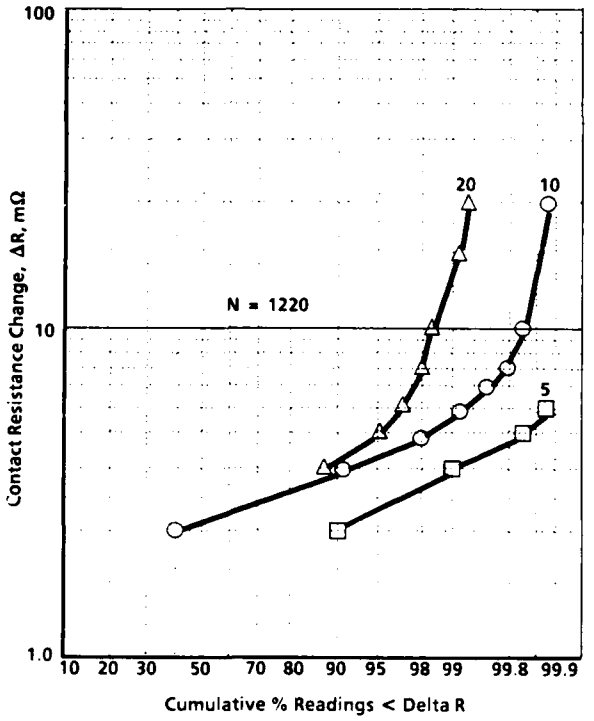


Figure 5. Performance of Mated Pin and Socket(25 DB) Connectors in Class II FMG Environment

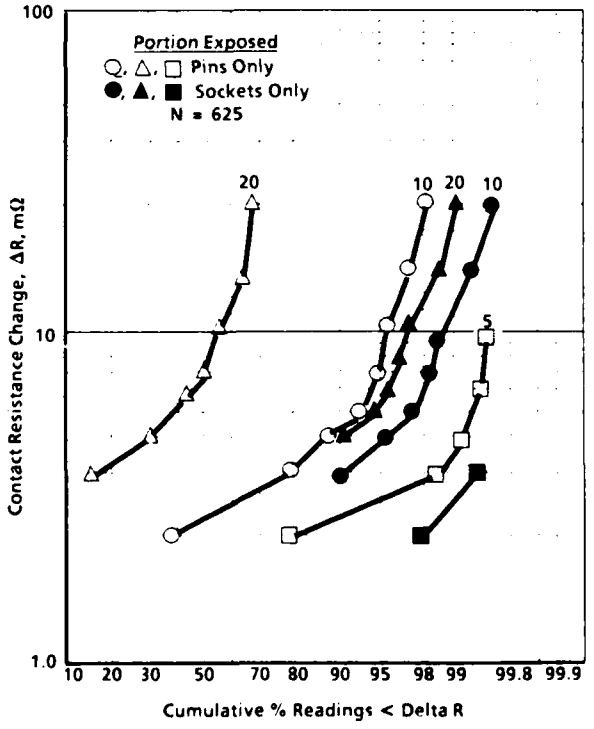


Figure 6. Performance of Unmated Pin and Socket (25 DB) Connectors in Class II FMG Environment

1 to 2 orders of magnitude higher than the SO_2 levels found in the average Class II field environment.

The experimental results are shown in Figure 7 for the two edge card connector systems. Again, the data show no measurable effect of SO_2.

Sulfide Environments. Additional data were obtained on the edge card systems in four environments in which only H_2S and/or SO_2 were the primary reactive pollutants. These results are summarized in Figure 8 for performance of the mated connector systems. These data support the primary conclusion reached from Figure 2. This is that exclusive sulfide environments even at concentrations 1 to 2 orders of magnitude above field levels are essentially benign. At most, they do not produce useful acceleration relative to field experience.

STATISTICAL IMPLICATIONS

The results in Figure 8 reflect a concern which has been expressed over the use of these newer testing methods. Figure 8 clearly shows that the FMG techniques will (1) produce higher rates of degradation, and (2) unacceptably high failure rates on specific designs and materials in comparison to historically used tests. The concepts of acceptable or unacceptable must, however, be related to the specific application as well as the failure criterion.

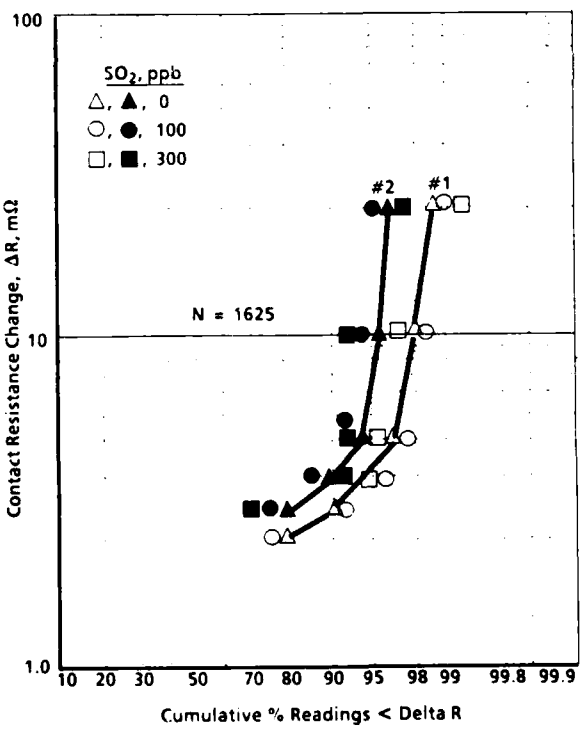


Figure 7. Effects of SO_2 Additions to Class II on Edge Card Statistics 20 Day Exposures

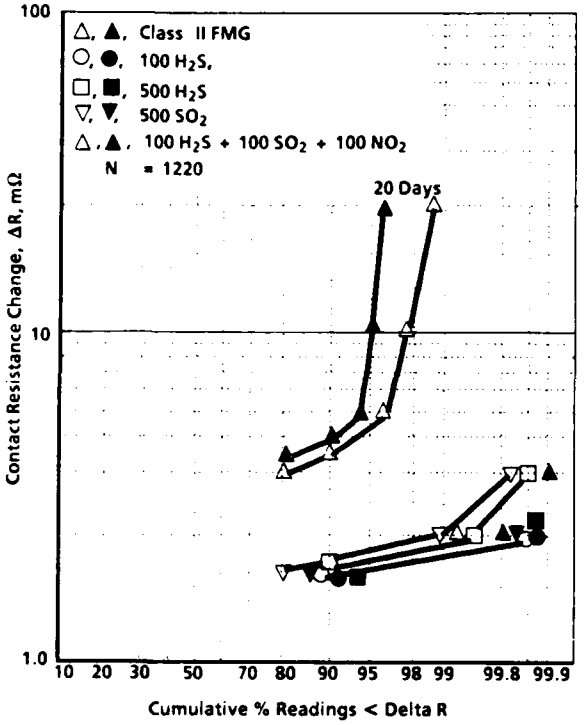


Figure 8. Edge Card Connector Performance in Dilute, Sulfide-Only Environments; 20 Day Exposures, 30°C, 70% RH

The results in Figures 3 through 6 show the results of actual measurements for exposure times up to 20 days. Within this period, two features should be noted in the data. First, there were relatively few high-resistance observations as found in the "tails" of the distributions. Second, the majority of these readings were typically no more than about 20 milliohms change. However, for short exposure times typical resistance changes which could be observed were far less than 10 milliohms.

The data used as examples in this paper were intentionally taken from some of the poorest performing connectors based on laboratory experience. Many systems will exhibit distributions shifted far to the right of those in Figures 3 through 6 ($\ll$0.1% degradation).

These observations are significant with respect to minimum samples sizes required for testing of this type, if an objective is either testing to failure or to observe sufficient change for comparisons. For example, if an objective might be testing to a 10 milliohm failure criterion, the results in Figure 3 suggest that a sample size in excess of 100 would be required to observe even a single failure. Experience gained on other connector systems, however, indicate that minimum sample sizes more than a factor of 10 larger are actually required to meet the same objective. In other words, the use of small sample sizes is not consistent with the typical degradation statistics found in the FMG environments.

Similar conclusions apply to the question of minimum sample size required to support modern reliability requirements. These might range, for example, from $<$0.001% per 1000 hours in "high-end" systems to $>$0.01% in some consumer electronics. In the first example, a sample size in excess of 3000 might be required to support that failure rate at a 95 percent confidence limit. Conversely, for the second example, a sample size of about 600 might suffice.

Table 2 shows the results of calculations on projected failure rates for the data shown in Figures 3 and 5. Values are shown for the upper and lower bounds of the 95 percent confidence intervals. Calculations were also made for failure criterion of 10 and 20 milliohms ΔR. Specific decisions regarding these data would have to be application dependent. However, as a general statement, the failure rates implied by these data would be inadequate for many high-reliability applications even though the percentage of contacts actually observed to degrade during test was relatively small.

TABLE 2. PROJECT FAILURE RATES OF MATED CONNECTORS IN CLASS II FMG ENVIRONMENTS

	Estimated Failure Rates, %/KOH[a]			
	R = 10m		R = 20m	
Connector	Upper	Lower	Upper	Lower
Edge Connector #1	.030	.021	.018	.011
Edge Connector #2	.060	.047	.047	.035
Pin and Socket	.026	.012	.015	.004

(a) KOH = 1000 Field Operating Hours @ 95% Confidence Limits

CONCLUSIONS

This paper is an extension of earlier studies which described the degradation of materials surfaces in coupon form exposed to various FMG environments. The current studies have shown that mated connector products may degrade even under Class II FMG conditions. However, observed rates are typically many orders of magnitude lower than rates found on coupons. This behavior has been attributed to the effects of environmental attenuation/ shielding by connector housings, as well as other aspects of mechanical design.

An important finding which is consistent with earlier coupon data concerns the sulfide-chloride interactions. All data indicate that exclusive sulfide environments are relatively benign and produce neither useful acceleration nor realistic reaction mechanisms in comparison to the field. These results confirm that proper simulation of these interactions must be used for realistic product evaluation.

Results from several types of commercial connector hardware have shown that the typical rate of hardware degradation in the Class II FMG environment is quite low. However, failure rates will typically be far higher than those expected for H_2S or SO_2 exposures. The practical implications of this observation are (1) such tests cannot be conducted quickly (>5 to 10 days), and (2) large samples sizes (>500 to 1000 pin count) must be used, particularly if a test objective is to achieve measurable degradation such as would be required for comparative evaluations. In practice, however, test duration should be matched to both the required product life and expected test acceleration factors.

REFERENCES

1. Abbott, W. H., "The Development and Performance Characteristics of Mixed Flow Gas Test Environments", Proc. 33rd IEEE/Holm Conference on Electrical Contacts, pp 63 (1987).

2. Abbott, W. H., "Corrosion of Porous Gold Plating in Field and Laboratory Environments", Plating and Surface Finishing, pp 72, (November, 1987).

3. ISA Specification ISA-d571.04

4. Campbell, W. E. and Thomas, U. B., "Tarnishing Studies on Contact Materials" Proc. of the 34th Intl. Res. Symp. on Electrical Contact Phenomena, Orono, pp 83, (1966).

5. Wiltshire, B., "A Survey of Contact Atmospheres in U.K. Telephone Exchanges", British Telecom Technical Journal, 2, 1, pp 74, (January, 1984).

6. Mollen, J. C. and Trzeciak, M., "Analysis of Copper Films After Field Exposures", Electrical Contacts, pp 37 (1970).

7. Rice, D. W., et al., "Atmospheric Corrosion of Nickel", J. of the Electrochem. Soc., 127, pp 563 (1980).

8. Antler, M., "Field Studies of Contact Materials in Telephone Central Offices", Proc. of the 11th Intl. Conf. on Electrical Contact Phenomena, pp 197, West Berlin (1982).

9. Hacque, C. A. and Antler, M., "Atmospheric Corrosion of Clad Palladium-Silver Alloys - Part I: Film Growth and Contamination Effects", Electrical Contacts, pp 183 (1981).

FLEXIBLE FLUOROPOLYMER LAMINATES

Paul J. Martin, Melvin P. Zussman, Craig S. McEwen, James Wright
E. I. du Pont de Nemours & Co., Inc.
Wilmington, Delaware, USA

ABSTRACT

Low dielectric-constant printed-wiring-boards fabricated from flexible fluoropolymer laminates offer fast signal propagation to designers of high-speed data processing systems. The composition of two such flexible fluoropolymer laminates are described. They are characterized, and compared to a conventional flexible laminate based on polyimide with an acrylic adhesive.

Flexible printed-wiring-board (PWB) laminates literally add a whole new dimension to circuit design. These laminates offer fabricators the option of bending PWB's into 3-D shapes to fit confined spaces, or of using the PWB itself as a connecting cable. In general, "high-reliability, high-performance" flex circuits (eg. circuits for the military) are fabricated from laminates utilizing a polyimide film reinforcement with an acrylic adhesive and rolled-annealed copper.[1] Such a laminate is illustrated in Figure 1.

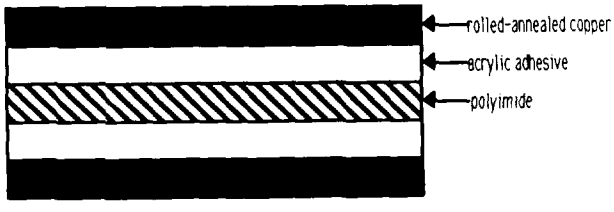


FIGURE 1: POLYIMIDE/ACRYLIC ADHESIVE FLEX LAMINATE

Higher clock rates of IC chips are driving system designers to low dielectric constant (K) materials. Even though the polyimide/acrylic adhesive flexible laminate exhibits a typical K around 3.6, there is a need for a flexible laminate with an even lower dielectric constant. This low-K flex laminate could be used in such applications as back-plane connectors, probe cables for high-speed test equipment, high-speed delay lines for oscilloscopes, microwave boards, and microwave antennae. This paper describes the development of low-K flex laminates.

When searching for resins/adhesives to fabricate low-K flex laminates, one naturally turns to fluorinated thermoplastics.[2] These materials offer low K with several other attractive properties, such as low dissipation factor, low moisture absorption, excellent chemical resistance, and thermal stability.

The choice of reinforcement is dictated by the combined requirements of low K and flexibility. The reinforcement must also withstand fluoropolymer processing temperatures of up to 360°C. Polyimide film is one choice. Alternatively, a woven polyaramid fabric can be employed as a reinforcement.

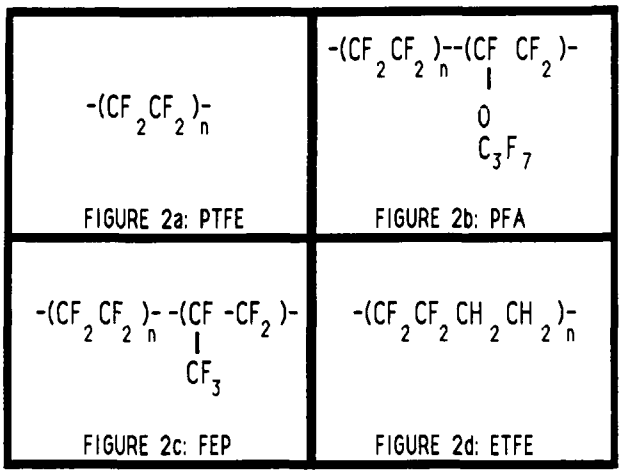
MATERIALS

POLYTETRAFLUOROETHYLENE (PTFE)[3] - PTFE, illustrated in Figure 2a, is a homopolymer of tetrafluoroethylene (TFE). PTFE is coated from a 60% solids aqueous dispersion containing a small amount of wetting agent. Selected properties of PTFE are given in Table 1.[4,5]

TETRAFLUOROETHYLENE/PERFLUOROPROPYLENE VINYL ETHER COPOLYMER (PFA)[3] - PFA, illustrated in Figure 2b, is a random copolymer of TFE and about 1 mole% perfluoropropylenevinylether. PFA is utilized as a 0.5-mil extruded sheet. Selected properties of PFA are given in Table 1.[6]

TETRAFLUOROETHYLENE/HEXAFLUOROPROPYLENE COPOLYMER (FEP)[3] - FEP, illustrated in Figure 2c, is a random copolymer of TFE and about 6 mole % hexafluoropropylene. FEP is utilized as an extruded sheet. Selected properties of FEP are given in Table 1.[4,5]

TETRAFLUOROETHYLENE/ETHYLENE COPOLYMER (ETFE)[3] - ETFE, illustrated in Figure 2d, is a largely alternating copolymer of TFE and ethylene. ETFE is utilized as an extruded sheet. Selected properties of ETFE are given in Table 1.[6]



POLY(PYROMELLITIMIDO-4,4'-OXYDIPHENYLENE) (PMDAODA) - PMDAODA, illustrated in Figure 3, is a polyimide film cast from a polar solvent as polyamic acid and subsequently imidized. Selected film properties are given in Table 1.[7,8]

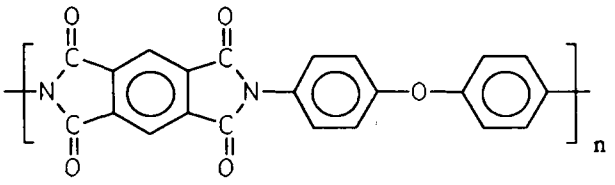


FIGURE 3: PMDAODA POLYIMIDE

POLY(*p* - PHENYLENE TEREPHTHALAMIDE) (PPTA) - PPTA , illustrated in Figure 4, is a lyotropic polyaramid. The resin is wet spun into fibers which are used to prepare 55 denier yarn. The yarn is woven into a plain weave fabric with 60 ends/inch and a basis weight of 1 oz/square yard. Selected fiber properties are given in Table 1.[9,10]

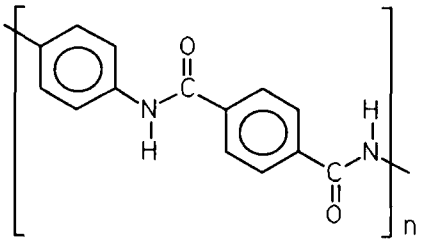


FIGURE 4: PPTA POLYARAMID

ROLLED-ANNEALED COPPER FOIL (RA COPPER) - RA Copper is prepared from electrolytic cathode copper with an oxygen content less than 0.03%. The foil is 0.0014" thick with a weight of 1 oz./sq. yard. The foil has a tensile modulus of 17,000 kpsi and a coefficient of thermal expansion (CTE) of 17.7 ppm/°C.[11] Adhesion of the copper to resins is enhanced by means of chemical treatment to one foil surface.

TABLE 1: SELECTED PROPERTIES OF LAMINATE MATERIALS

Property	Units	PTFE	PFA	FEP	ETFE	PMDAODA	PPTA
Max Use Temp.	°C	260	260	205	>150	400	425
Melting Point	°C	327	305	290	270	none	none
Tensile modulus	Kpsi	76	62	75	120	430	18,000/1000[a]
Coeff.Therm. Exp.	ppm/°C	151	184	182	113	31	-5.2/41.4[a]
Specific Gravity	g/cc	2.2	2.15	2.15	1.7	1.42	1.44
Flammability	(UL94)	V0	V0	V0	V0	V0	V0
Dielectric Constant	(1 MHz)	2.05	2.05	2.1	2.6	3.4	3.8[b]
Dissipation Factor	(1 MHz)	.00009	.0002	.0004	.005	.01	

a. Values are listed for PPTA fiber in the axial/radial direction.

b. The dielectric constant for PPTA fiber was calculated from measurements of composite values

POLYIMIDE-REINFORCED LAMINATE

LAMINATE - One means of fabricating a low-K flex laminate is to replace the acrylic adhesive of a conventional flex laminate (Figure 1) with a fluoropolymer. The advantage of this approach is that the resulting low-K laminate would process into a PWB almost identically to the conventional flex laminate. To achieve this replacement of acrylic adhesive by fluoropolymer, the adhesion at two interfaces must be satisfactory. The fluoropolymer must adhere to the polyimide, and, in turn, the copper must adhere the fluoropolymer. PFA, FEP, and ETFE films will all stick to PMDAODA film by means of a surface treatment of the PMDAODA followed by hot-press lamination of the fluoropolymer. Furthermore, copper foil with the proper surface treatment will hot-press laminate to the fluoropolymer. PFA is the preferred fluoropolymer because it combines low dielectric constant with a high continuous-use temperature (260°C). High use -temperature is important as the laminate must subsequently withstand such processes as soldering (288°C) and bonding a protective cover-coat (290°C).

The replacement of acrylic adhesive with PFA produces some unanticipated advantages. Moisture absorption drops from ca. 3% by weight to substantially less than 1%. This results in less variation in electrical properties with humidity. Impedence control is better due to near-zero flow of the fluoropolymer during circuit fabrication, a result of the high melting point of PFA. The laminate is flame retardant. Furthermore, there is some indication of improved resolderability/ repairability of a PFA laminate, again because the fluoropolymer does not melt at soldering temperatures.

A flex laminate with 0.5-mil PFA laminated to 1-mil PMDAODA is illustrated in Figure 5. Typical properties of the PMDAODA/PFA laminate are compared to those of a 1-mil polyimide/1-mil acrylic adhesive laminate in Table 2.

FIGURE 5: PMDAODA/PFA FLEX LAMINATE

COVERCOAT - Once circuit lines are etched, it is generally necessary to protect them with a covercoat. The acrylic adhesive can be hot-press laminated as the covercoat. Either ETFE or FEP can be used if a lower dielectric constant is desired. These materials require a higher lamination temperature than the acrcylic adhesive (290°C vs. 190°C), but this is still below the melting point of PFA.

POLYARAMID-REINFORCED LAMINATE

LAMINATE - A flex laminate with even lower dielectric constant and better dimensional stability can be made with a PPTA fabric reinforcement. PPTA fabric exhibits textile-like flexibility, and also has a lower CTE and higher modulus than PMDAODA or even conventional woven fiberglass reinforcement. This allows the fabrication of a flex laminate with excellent dimensional stability. Furthermore, only a small amount of reinforcement is required, so the overall dielectric constant of the flex laminate can be quite close to that of the fluoropolymer resin. The PPTA/fluoropolymer composite is produced by dip-coating the polyaramid fiber in the PTFE dispersion, drying to remove water, then sintering the fluoropolymer at 360°C. The laminate is completed by hot-pressing treated copper foil onto the composite at 360°C and 200 psi.

Improvements in dielectric constant and dimensional stability over the PMDAODA/PFA laminate are made at the sacrifice of some processibility. In general, PTFE surfaces must be chemically etched to promote the adhesion of plated metals or covercoats.

A 5-mil flex PPTA/PTFE laminate is illustrated in Figure 6. Typical properties of the laminate are given in Table 2.

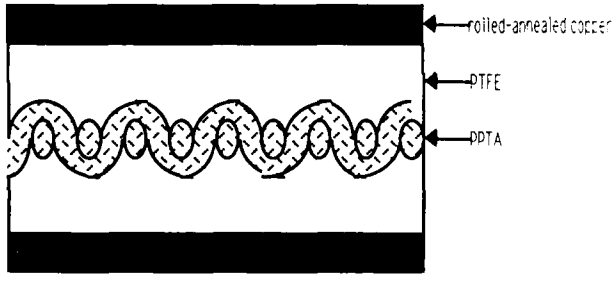


FIGURE 6: PPTA/PTFE FLEX LAMINATE

COVERCOAT - Covercoats which are suitable for the PMDAODA/PFA flex laminate are also satisfactory for the PPTA/PTFE laminate. As noted previously, the surface of the PTFE may have to be etched to assure good adhesion of the covercoat.

SUMMARY AND CONCLUSIONS

Typical properties of a conventional 1-mil polyimide/ 1-mil acrylic adhesive flex laminate, a PMDAODA/PFA flex laminate, and a PPTA/PTFE flex laminate are summarized in Table 2.

TABLE 2: TYPICAL PROPERTIES OF LAMINATES

	polyimide/ acrylic	PMDAODA/ PFA	PPTA/ PTFE
Electrical			
dielectric constant (at 1 MHz)	3.5	2.6	2.2
propagation delay (ns/ft)	1.90	1.64	1.51
dissipation factor	0.02	0.01	0.003
dielectric strength (V/mil)	4000	4000	1300
volume resistivity (M-Ω-cm)	10^8	10^{10}	10^{17}
surface resistivity (M-Ω)	10^7	10^7	10^{15}
Physical			
dimensional stability (%)	0.15	0.6	0.10
peel strength (pli)	12.0	10.0	12.0
solder float (288°C for 10 sec)	OK	OK	OK
chemical resistance	good	excellent	excellent
flammability	supports flame	UL-V0	UL-V0
moisture absorption (%)	2.5 -3.5	<1.0	-

Time Domain Reflectometry measurements from 500 MHz to 10 GHz indicate that the mean dielectric constants of PMDAODA/PFA and PPTA/PTFE are 2.6 and 2.2, repectively. The mean dissipation factors are 0.015 and 0.005, respectively.

In conclusion, it appears that flexible fluoropolymer laminates are a viable alternative for the high-speed-system designer. If a lower dielectric constant is desired without giving up ease of processibility, a PMDAODA/PFA laminate can be utilized. If an even lower dielectric constant is needed, a PPTA/PTFE laminate is required.

Flexible fluoropolymer laminates offer the additional advantages of low moisture absorption, flame retardance, good thermal stability, excellent chemical resistance and, in the case of PPTA reinforced laminates, superior dimensional stability.

REFERENCES

1. S. Gurley, "Flexible Circuits", p. 3, Marcel Dekker, Inc., New York (1984)
2. J. R. Cannizzaro, "Printed Circuit Handbook", (C. F. Coombs, Editor), p. 25-2, McGraw-Hill Book Company, New York (1979)
3. D. C. England, R. E. Uschold, H. Starkweather and R. Pariser, Proceedings of *The Robert A. Welch Conferences on Chemical Research* XXVI. Synthetic Polymers, Houston, Texas 203-15 (1982)
4. S. V. Gangal, "Kirk-Othmer Encyclopedia of Chemical Technology", 3rd Ed., 11, p. 1-35, John Wiley & Sons, New York (1980)
5. G. P. Koo, "Fluoropolymers" (L. A. Wall, Editor), p. 507-41, John Wiley & Sons, New York (1980)
6. R. L. Johnson, "Kirk-Othmer Encyclopedia of Chemical Technology", 3rd Ed., 11, p. 35-46, John Wiley & Sons, New York (1980)
7. "Kapton® Polyimide Film: Summary of Properties", E. I. Du Pont de Nemours, Inc. (1987)
8. F. A. Bystry-King and J. J. King, "Engineering Thermoplastics: Properties and Applications" (J. M. Margolis, Editor), p. 315-33, Marcel Dekker, Inc., New York (1985)
9. "Data Manual for Kevlar®49 Aramid", E. I. Du Pont De Nemours, Inc. (1986)
10. M. G. Dobb, " Strong Fibers" (W. Watt and B. V. Perov, Editors), p. 673-704, Elsevier Science Publishers B.V., Amsterdam (1985)
11. "Wrought Copper Foil for Printed Circuits", Olin Corporation (1988)

THE CHEMISTRY OF RESIST MATERIALS

Scott A. MacDonald
IBM Research Division
Almaden Research Center,
San Jose, California 95120, USA

Organic imaging materials (commonly called resists) play an important role in the manufacturing of semiconductor devices. However, the common resist materials used today were not specifically designed for this application. These polymeric materials were originally developed for printing legible text and were simply adopted by the electronics industry with only minor modifications. Fortunately, the chemistry of these systems is readily extendable to smaller geometries. The microelectronic manufacturing process today is operating at a resolution of 1 micron which allows efficient production of, for example, 1 Megabit DRAMS and is rapidly moving toward production of 4 Megabit DRAMS with submicron geometries (1). Several excellent, general reviews that discuss both the chemistry of resist systems, as well as the use of these materials to fabricate microelectronic devices, have been published recently (2-4).

Traditionally, one could cleanly classify a resist system as either "positive tone" or "negative tone", depending on its response to light. Recently, that distinction has blurred for some systems. For example, it is possible to modify the chemistry of diazonaphthoquinone/novolac resin (a standard positive tone resist) such that it will function as a high resolution, negative-tone system (5-7). In this case, the image reversal is accomplished by adding a base catalyst (such as ammonia or an alkylamine) to the film and introducing two additional processing steps. Recently, we have developed a resist system that will function in either a positive or a negative tone, depending on the wavelength of the exposing radiation (8). This system yields a positive image when exposed to 400 nm light, but yields a negative tone image when exposed to 313 nm light. By designing a photomask that contains regions transmitting 400 nm light (but not 313 nm light), areas transmitting 313 nm light, and opaque regions, one can define three regions within the film in a single exposure.

The previously described resist systems, as well as the ones commonly found in manufacturing, all use some sort of "wet" development step to obtain the relief image. This developer is traditionally an organic solvent or an aqueous base solution. Several research groups are investigating schemes to use plasma or reactive ion etching conditions, to develop the relief image. One approach involves treating the photoresist (after exposure) with an organometallic reagent such that an inorganic component (like silicon) is selectively incorporated into either the exposed, or the unexposed regions of the film. When placed into an plasma environment the regions without the inorganic additive, will be etched to substrate. Coopmans and Roland have described a system in which a diazonaphthoquinone/novolac resin resist is exposed to UV light and then treated with a silylating agent (9). In this process the silicon diffuses into the exposed areas, generating a negative tone resist. The mechanism and kinetics of this system have been explored by a group at Philips (10). IBM has described a system in which the UV exposure generates chemically reactive sites which subsequently form a covalent bond to the silylating agent (11). The kinetics and mechanism of this chemistry have been studied by the Kodak Research group (12).

REFERENCES

1. P.H. Sinzer, "4M-DRAM Technology-IBM Lead the U.S", Semiconductor Intl. 1987, 10(6), P.50.

2. C. G. Willson, "Organic Resist Materials-Theory and Chemistry", in "Introduction to Microlithography": L. F. Thompson, C. G. Willson, M. J. Bowden, Eds; ACS SYMPOSIUM SERIES #219, American Chem. Soc.,; Washington D.C., 1983; p.87.

3. M. J. Bowden, "A Perspective on Resist Materials for Fine Line Lithography" in "Materials for Microlithography"; L. F. Thompson, C. G. Willson, J.M.J. Frechet, Eds.; ACS SYMPOSIUM SERIES #266, American Chem. Soc.,; Washington D. C.,1984; p.39.

4. W. M. Moreau, "Semiconductor Lithography, Principles, Practices and Materials" Plenum Press, New York, New York, 1988.

5. S. A. MacDonald, H. Ito, and C. G. Willson, Microelectronic Engineering, 1, 1983, p.269.

6. H. Moritz, IEEE Trans. Elect. Dev., ED-32(3), 1985, p.672.

7. R. M. R. Gijsen, H. J. J. Kroon, F. A. Vollenbroek, and R. Vervoordeldonk, "A Quantitative Assessment Of Image Reversal, A Candidate For A Submicron Process With Improved Linewidth Control", Advances in Resist Technology and Processing III, C. G. Willson, Editor, Proc. SPIE 631, 1986, p. 108.

8. W. D. Hinsberg, S. A. MacDonald, L. A. Pederson, and C. G. Willson, "Zero-misalignment lithographic process using a novel photoresist with wavelength-selected tone", Advances in Resist Technology and Processing V, S. A. MacDonald Editor, Proc. SPIE 920, 1988.

9. F. Coopmans and B. Roland, Solid State Technology, 30(6), 1987, p.93.

10. R. J. Visser, J P. W. Shellekens, M. E. Renhman-Huisken, J. Van Ijzemdpprn, "Mechanism and Kinetics of Silylation of Resist Films From The Gas Phase", Advances in Resist Technology and Processing IV, M. J. Bowden Editor, Proc. SPIE 771, 1987, p.111.

11. S. A. MacDonald, H. Ito, H. Hiraoka, and C. G. Willson, Proc. SPE Regional, Technical Conference, Ellenville, New York, 1985, p.177.

12. T. B. Burst, S. R. Turner, "Mechanistic Study Of A Selective Gas Phase Reactive Microlithographic System", Advances in Resist Technology and Processing, M. J. Bowden, Editor, Proc. SPIE 771, 1987, p.102.

ELECTRICAL PROPERTIES OF METAL FILLED COMPOSITES

K. Andres, Y. S. Ho, J. Lodge, P. Sturman, J. J. Kramer
University of Delaware
Newark, Delaware 19716, USA

ABSTRACT

The electrical properties (permittivity at X band and D.C. conductivity) of a series of composites consisting of metal particles, up to 35 volume percent, in a non-conducting matrix were examined. For the metal filler, particulates (spherical-like particles) of Al, Cu, and Ni, of various particle size between 1μm and 200μm were used, as well as 8μm nickel-coated carbon fibers with different aspect ratios. The matrix was epoxy, polyester or wax.

An evaluation of the microstructures of the composites was made, with an image analyzer used to determine the extent of particle to particle contact and fiber orientation distribution. With the Ni both a random as well as a more ordered arrangement of the particulates was used. The latter was obtained by processing the composite in a magnetic field.

The complex permittivity of the composites was determined by scattering parameter techniques using a network analyzer; D.C. conductivity measurements were made with an electrometer. The dielectric results for the particulate samples do not seem to follow effective medium or mean field theories. There is evidence to to suggest that interfacial effects such as polarization and charge injection may be important. Percolation behavior was not observed. For the fiber samples the results can be analyzed based on a simple model that takes into account the orientation distribution of the fibers and their aspect ratio.

THE ELECTRICAL PROPERTIES OF COMPOSITES

are generally not a simple average of the properties of the two or more constituent phases. This is particularly evident for a system when one of the phases is highly conducting while the other is an insulator. For such composites the permittivity[1] and the electrical conductivity usually do not follow "rule of mixture" or "inverse rule of mixture" behavior. An exception is when the phases involved have planar geometries, in particular a laminated structure. For these structures the conductivity is expected to follow the rule of mixture for transport parallel to the laminations and the inverse rule of mixtures in the direction normal to the laminations. Even here, however, interfacial phenomena might affect the simple relationships when the scale of the microstructure is small. The rule of mixtures and the inverse rule of mixtures thus serve only as crude "upper" and "lower" bounds to the expected properties of any composite system.

In composite systems of metals embedded in a non-conducting polymer, the difference in conductivity between the two phases is as large as eighteen orders of magnitude. As a result the separation between the upper and lower bounds is correspondingly huge. At a particular volume fraction of embedded metal, the measured value of the electrical property will depend on which of the microstructural properties are important (particle shape, size, size distribution and geometrical arrangement within the system) in addition to the bulk material properties and any interfacial phenomena that may exist. To adequately deal with the microstructure both its characterization and then its correlation with electrical properties must be determined. This has been the subject of extensive investigations over the years beginning with the early work of Maxwell [1,2].

While significant progress has been made in characterizing systems and correlating the microstructure with electrical properties, considerable controversy still remains. Most of the work has been done with particulate systems (spherical-like particles) since one can avoid some of the difficulties that occur with rod or plate-like systems. For these latter systems particle orientation ef-

[1]For this paper the term permittivity will include both the real and imaginary parts. Dielectric constant will refer to the real part of the relative permittivity and the symbol ϵ' will be used. The term imaginary dielectric constant, ϵ'' will be used when needed to define the imaginary part of the relative permittivity.

fects leading to anisotropic behavior and particle contact effects must usually be dealt with. Experimentally these systems are difficult to synthesize with any degree of control. Also the extent of particle to particle contact and the orientation distribution are difficult to determine experimentally.

In particulate systems a number of expressions have been developed in an attempt to account for the variation in permittivity of the composite with particulate volume. Included are the Maxwell-Garnett [3]

$$\frac{\epsilon(\omega)-\epsilon_i(\omega)}{\epsilon(\omega)+2\epsilon_i(\omega)} = V_{fm}\frac{\epsilon_m(\omega)-\epsilon_i(\omega)}{\epsilon_m(\omega)+2\epsilon_i(\omega),} \qquad (1)$$

the Bruggeman [4]

$$V_{fm}\left[\frac{\epsilon_m(\omega)-\epsilon(\omega)}{\epsilon_m(\omega)+2\epsilon(\omega)}\right] = -(1-V_{fm})\left[\frac{\epsilon_i(\omega)-\epsilon(\omega)}{\epsilon_i(\omega)+2\epsilon(\omega)}\right] \qquad (2)$$

and the Landau, Lifshitz, Looyenga[5]

$$\epsilon(\omega)^{\frac{1}{3}} = (1-V_{fm})[\epsilon_i(\omega)]^{\frac{1}{3}} + V_{fm}[\epsilon_m(\omega)]^{\frac{1}{3}}. \qquad (3)$$

Here $\epsilon(\omega)$, $\epsilon_m(\omega)$ and $\epsilon_i(\omega)$ are the frequency dependent constants of the composite, metal and insulator respectively, and V_{fm} is the volume fraction of metal particulates in the composite.

The Maxwell-Garnett expression is a mean field expression. It can also be regarded as a first order multiple scattering evaluation of the propagation of electromagnetic radiation through a heterogeneous system in which the particulates are separated from one another. It has been successfully used to explain the resonances that occur in systems of metal particulates in a dielectric at optical frequencies [6]. In the metal rich region where particulate coalesence begins (i.e., extensive particle to particle contact leading to continuous chains -i.e., percolation) this theory breaks down. For $\epsilon_m(\omega)$ large, as is usually the case below optical frequencies, a singular point occurs at $V_{fm} = 1$, while experimentally a singularity is observed at the onset of percolation. This occurs at a much lower volume fraction.

The Bruggeman equation, an effective medium expression, suffers from the similar problem: a singularity occurs at $V_{fm} = 0.33$. In the Looyenga equation no unique singularity occurs at $V_{fm} = 1$, nor at any other volume fraction, but the equation requires the specific evaluation of the metal dielectric constant.

The above equations, while not useful in dealing with the phenomenon of percolation, could in principle be applied to explain data at volume fractions considerably below the onset of percolation. Although they have been written above in terms of the permittivity they express the appropriate form for any other transport property such as conductivity. Modifications can be made to deal with anisotropic behavior for rod or plate-like particles by considering the symmetry of the system and the properties of any second-rank tensor.

The onset of percolation in the metal particulate systems leads to dramatic changes in both the dielectric constant and the conductivity. These dramatic changes have been dealt with by the use of "percolation theory" which attempts to define properties in a narrow region, the critical regime, around the percolation threshold [7]. This threshold for a uniform particulate systems is a function of both the degree of disorder in the system and its dimensionality. For a completely disordered 3-D system percolation begins at V_{fm} of ~ 0.145 [8]; while for a simple cubic array the value is 0.311 [7].

In the critical regime, as one approaches the percolation threshold from lower metallic volume fractions, more particle to particle contact occurs, giving rise to larger clusters. This can be shown to lead to a divergence of the real part of the dielectric constant when the threshold is approached from tne insulating side and to a disappearing D.C. conductivity when approached from the metallic side. Appropriate relationships exist for the conductivity and the dielectric constant in terms of the correlation length ξ which corresponds to the average radius of a typical percolation cluster. Specifically

$$\xi \ \alpha \ \mid V_{fm} - V_{fmc} \mid^{-\nu}, \qquad (4)$$

where V_{fmc} is the volume fraction of the metal at percolation and ν has a value less than 1 for a three dimensional system.

Scaling laws have been developed to define the correct exponential relationship between the conductivity or dielectric constant and the correlation length, as well as the frequency dependence of these two transport properties. A few experimental measurements have been made in an attempt to verify the correct exponential relationships and the validity of the scaling laws [9,10,11]. In general these studies have only been partly successful.

Recently an attempt was made to develop a general transport expression which avoids dealing separately with both the percolation regime and the low volume fraction region of particulates [12]. This approach takes into account the extent of particle to particle contact and chain formation by defining an effective aspect ratio for the system. The aspect ratio is used to evaluate an average particulate shape factor $< h >$ which serves as a measure of the extent of particle to particle contact. The complex dielectric constant for a conducting particulate system can then be expressed as

$$\epsilon(\omega) = \epsilon_i(\omega)\left[\frac{(1-<h>)(1-V_{fm})+<h>V_{fm}}{(1-V_{fm})^2(1-<h>)}\right] \qquad (5)$$

For isolated particulates, as encountered when $V_{fm} \rightarrow 0$, $< h >= \frac{2}{3}$. As one approaches percolation $< h >$ should approach 1. This equation avoids the difficulties encountered by the Maxwell-Garnett and Bruggeman expres-

sions regarding the volume fraction at which the singularity occurs. Here it should be encountered at the percolation threshold. Compliance with equation 5 requires a careful experimental evaluation of the shape factor.

The development leading to equation 5 has been extended to deal with short fiber systems of various fiber aspect ratios [2]. For a dilute solution of aligned conducting fibers one obtains

$$\epsilon_1(\omega) = \epsilon_i(\omega) + \frac{\epsilon_i(\omega)V_{fm}}{Y_1} \tag{6}$$

$$Y_1 = [1 - h(a)] \rightarrow 0$$

for measurements along the fiber direction and

$$\epsilon_2(\omega) = \epsilon_3(\omega) = \epsilon_i(\omega) + \frac{\epsilon_i(\omega)V_{fm}}{Y_2} \tag{7}$$

$$Y_2 = \frac{h(a)}{2} \rightarrow \frac{1}{2}$$

for measurements normal to the fibers. The results for fiber systems with an orientation distribution can be evaluated by considering the effects of rotation from the principle direction on the properties of a second-rank tensor. In dilute fiber systems, where relatively little fiber to fiber interaction or contact occurs, it is expected that the contribution to the dieletric constant from individual fibers at different orientations can be summed.

For particulate systems, scaling effects related to particle size and particle size distribution have usually not been considered. In general, scaling effects are not important provided three conditions exist [13]: (1) The particulate size is small relative to the wavelength of the radiation used. (2) Interfacial effects such as interfacial polarization [14] and charge injection [15], tunneling [16] or hopping conductivity [17] do not occur and (3) that all electromagnetic properties of interest can be incorporated into a complex dielectric constant. For the systems presented here, condition (1) is satisfied and likely condition (3). If interfacial effects exist they are expected to be more pronounced for smaller particles. Likewise one might expect that particle size distribution will play a role in the onset of percolation.

EXPERIMENTAL MEASUREMENTS

This study of the electrical properties of composites was confined to D.C. conductivity measurements and microwave dielectric studies of different metal particulate filled non-conducting matrix composites, as well as conducting fiber filled composites. For the metal particulate composites the samples were synthesized by carefully blending together the metal powders (Cu, Ni or Al) with epoxy, polyester or wax and then subjecting the blended material to an appropriate curing cycle. For the copper and nickel, three different particle sizes were used, with the smallest close to 2 microns while the largest was 100 microns or greater. Special care was necessary with the larger particle sizes to prevent settlement during curing. For the fiber samples, the fiber orientation was established by controlling the flow characteristics of the composite into the mold. With one of the nickel-filled samples, some degree of particulate ordering was obtained by subjecting the blended composite to a magnetic field prior to curing.

The permittivity was determined from scattering parameter techniques made within the region of X band (8.5 - 11.5 GHz). For these studies, a Hewlett-Packard Network Analyzer (HP 8410) was used. The sample was appropriately shaped to fit into a waveguide and the TE_{10} mode was used. Thus, anisotropic behavior of the thin samples could be studied for both the fiber samples and the field treated nickel sample. To obtain the best accuracy, the thickness of the sample was made to be approximately one-quarter wavelength.

Measurements of the D.C. conductivity of the composites were made using a Keithley Model 617 Electrometer. To avoid anomalous measurements from possible surface currents, a guard ring was used. The samples were not electrically "conditioned" by subjecting them to a high voltage prior to the conductivity measurements [18]. All measurements were made at fields less than about 10 volts/cm to avoid electrical breakdown, particularly for the higher particulate volume fraction composites.

A Cambridge Instruments Quantimet 970 Image Analysis System was used in the determination of particulate size distribution and the extent of particle to particle contact in the composite. Fiber orientation distributions were determined optically on thin sections.

RESULTS & DISCUSSION

METAL PARTICULATE COMPOSITES - The data shown in Fig.1 and 2 are the real parts of the dielectric constants measured at X band for various volume fractions of metal filled particulate composites of Ni in wax, Cu in polyester and epoxy, and Al in polyester. The results represent an average of measurements taken over the frequency range 8.5 - 11.5GHz at 0.5 GHz intervals. For both Ni and Cu the impact of particle size, particularly for the small particles, is evident. Data on small silver particulates, reported earlier, is also consistent with these results [19]. The size distributions for some of the metal particles are shown in Table 1. For the smaller particles, blending difficulties limited the fabrication of high volume fractions. With the larger particles, the volume fraction was anticipated to extend to the percolation threshold and was beyond.

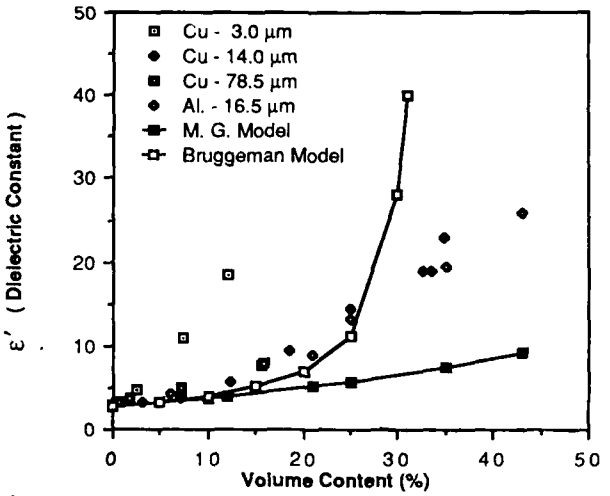


Fig. 1 - Dielectric Constant at 8.5 - 11.5GHz vs Vol % Cu (3 sizes) in Epoxy and Al in Polyester. Comparison with the Maxwell-Garnett and Bruggeman Models.

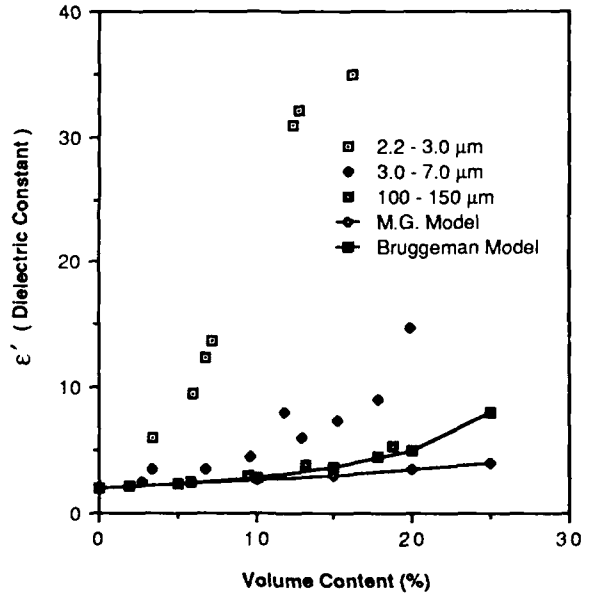


Fig. 2 - Dielectric Constant at 8.5 - 11.5GHz vs Vol % Ni (3 sizes) in Wax. Comparison with Maxwell-Garnett and Bruggeman Models.

Included in Fig. 1 and 2 are the expected values based on the Maxwell-Garnett and Bruggeman models. Clearly these models do not fit at the higher volume fraction, close to expected percolation values, but are also poor approximations even at low volume fractions. Since both models do not involve particle size effects their application is further in question.

The Looyenga expression can be made to fit the data, if an appropriate value for the dielectric constant of the metal is first calculated from the experimental data. For the 78.5 micron size copper particulates, a value of 134 is obtained. A subsequent comparison with the measured data is shown in Fig.3. While a reasonably good fit can be obtained for the 78.5 micron data,the dielectric constant of Cu is low; also the results for the 3.0 micron data require further consideration. Recently it has been suggested that particle size distribution should be incorporated into the Looyenga expression [20]. An analysis of this effect will be made in the future.

In a previous paper an attempt was made to analyze the measured results based on consideration of the aspect ratio as used in equation 5 [19]. For these studies, even at high particulate volume fractions, the aspect ratio never exceeded 3.3. This corresponds to a value of $< h >$ equal to about 0.9 and a value of the dielectric constant less than 20. While the use of equation 5 and the measured aspect ratio data provided an appropriate fit to the dielectric constant for samples with larger particulates, it was exceedingly poor for the smaller particulate samples of Cu, Ag and Ni. Difficulties, however, are encountered in trying to accurately determine the extent of 3-D particle contact from observations on a 2-D surface.

Table 1
Selected Particle Size Data

Cu(78.6μm)		Al(14.1μm)		Ni(2.6μm)	
Size(μm)	#Percent	Size(μm)	#Percent	Size(μm)	#Percent
<25	4	0 - 2.5	35	0 - 1.5	18
25 - 35	14	2.5 - 7.5	10	1.5 - 2.5	35
35 - 45	2	7.5 - 15	12	2.5 - 3.5	27
55 - 65	8	15 - 25	17	3.5 - 4.5	14
65 - 75	18	25 - 35	11	4.5 - 5.5	1
75 - 85	9	35 - 45	11	5.5 - 6.5	4
85 - 105	18	45 - 55	1	6.5 - 7.5	1
105 - 115	7	55 - 75	1		
115 - 125	14	85 - 105	2		
125 - 145	4				
>145	2				

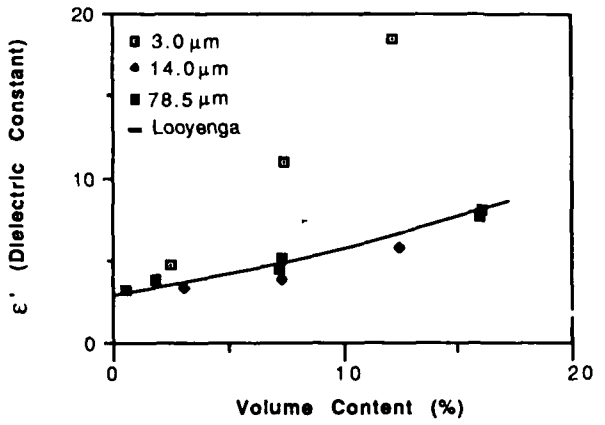


Fig. 3 - Dielectric Constant at 8.5 - 11.5GHz vs Vol% Cu (3 sizes) in Epoxy. Comparison with the Looyenga Model.

In the data of Fig.1 and 2 the dielectric constant shows a dramatic increase with V_{fm}, even below 20 volume percent for the fine particulate composites. Values are obtained that are not reached even at 40 volume percent for the larger particulate samples. As noted earlier, percolation in 3-D disordered systems could be expected at volume fractions as low as 0.145. Because the particles are not monodisperse this threshold probably occurs at a somewhat higher value. From purely geometrical arguments there is no reason to expect that percolation should be particle size dependent. However the average particle separation is considerably less for small particles at any given volume fraction, and the total surface area is greater. Both of these factors become important if interfacial effects influence the dielectric properties. At this point the results thus suggest that charge injection, interfacial polarization, hopping conductivity, or tunneling phenomena should be considered.

Measurement of the D.C. conductivity of the particulate samples did not show any increased conductivity for the finer particulate samples. In all cases, even for the highest values of V_{fm}, the conductivity was less than 10^{-12}(ohm cm)$^{-1}$. These measurements were on samples that had not been "conditioned" at a higher voltage exposure prior to the conductivity measurements. Such conditioning increases the conductivity, presumably by opening up conducting paths between closely separated particles, either by dielectric breakdown of the intervening polymer or by metal sputtering between particles, both due to high electric fields [10,18]. It has been suggested that polymer wetting of the metal particles could prevent good electrical contact even at high values of V_{fm}[18]. The dielectric measurements similarly were performed on samples that had not been conditioned.

These D.C. conductivity observations seem to rule out the possibility that appreciable tunneling or hopping conductivity is occurring, in contrast to observations elsewhere [21,22]. Thus the possibility of interfacial polarization and/or charge injection seems likely. Charge injection of electrons from the metal into the polymer matrix could lead to enhanced conduction if the injected region at the interface is sufficiently thick (i.e. greater than one half the average separation distance between particles). For a simple cubic array of 2μm size particles at a volume fraction of metal of 0.15 this injected region would have to exceed 0.5μm for increased conductivity to occur. Even without enhanced conductivity an impact of charge injection on the polarization phenomena and thus the dielectric constant is suggested. Interfacial effects are likely also to be influenced by the fact that polymer curing at the polymer-metal interface seems to be retarded by the presence of the metal [23].

The imaginary part of the dielectric constant is a measure of the effective A.C. conductivity of the system, in essence the electrical energy loss of the system. It includes, in addition to D.C. conductivity, all frequency dependent in-phase (with the electric field) current flow. With the exception of those samples where the real part of the dielectric constant is large, and shows a dramatic increase with V_{fm}, the imaginary part of the dielectric constant was small. For most cases tan δ, the ratio of the imaginary to the real part of the dielectric constant, was less than 0.1, the lower limit of sensitivity of the measuring system. Where the higher imaginary dielectric constants were found (accompanied by high real values), tan δ values as large as 1.0 were measured. This is not inconsistent with greater charge displacement that might occur with greater interfacial polarization or charge injection.

Some of the concepts of percolation theory can be examined in more detail to explain the results for those samples that show a dramatic increase in the dielectric constant with V_{fm}, but not an increased conductivity. It was noted earlier, that as the percolation point is approached from lower volume fraction of particulate, the dielectric constant is expected to show a rapid increase. Only at the percolation point and beyond does the resistivity begin to disappear. Thus true percolation may not have been reached. However, the lack of high conductivity at volume fractions of particulate in excess of the expected percolation point suggests the existence of current barriers at the points of apparent particle to particle contact. With this condition the properties of the cluster and the correlation length must be modified. With the exception of some thin film barrier, the geometry of the cluster will be the same but its correlation length will be decreased. In percolation theory, in the limit of zero frequency, the dielectric constant is shown to be related to the correlation length. However at increased frequencies the dielectric constant more specifically results from two effects: polarization between clusters and anomalous diffusion within a cluster [9]. This latter effect is more strongly dependent on the correlation length. Some indication which is the predominant factor can be obtained from measurements of the frequency dependence of the dielectric constant. The following functional relationship exists [11].

$$\epsilon'(\omega) \propto \omega^{-y} \qquad (8),$$

where $\epsilon'(\omega)$ is the real part of the dielectric constant. If comparison is made between the dielectric constants reported in an earlier paper [19] at 1 KHz and 10 GHz for Al in polyester (at the largest value of V_{fm}) a value of $y < 0.01$ is obtained. (The overall results for Al are similar to those seen here for Cu and Ni.) This compares with a value of 0.12 reported by other investigators [11]. A low value of y is more consistent with polarization between clusters rather than anomalous diffusion within a cluster as the source of the high dielectric constant. This is expected here if conducting barriers exist at contacting particle interfaces.

FIBER FILLED AND ALIGNED PARTICULATE COMPOSITES - The dramatic impact of fibers with a high aspect ratio on the dielectric properties of a composite, even at low volume fractions, is readily evident. The measured dielectric constant for nickel-coated fibers (8 μm diameter, aspect ratio 80) in epoxy ($V_{fm} = 0.0025$) is shown in Fig.4. For these studies an attempt was made to align the fibers by control of the flow conditions during molding. Measurements are shown over the frequency range 8.5 - 11.5 GHz for the electric field of the TE_{10} wave-guide mode both parallel and perpendicular to the general direction of the fiber.

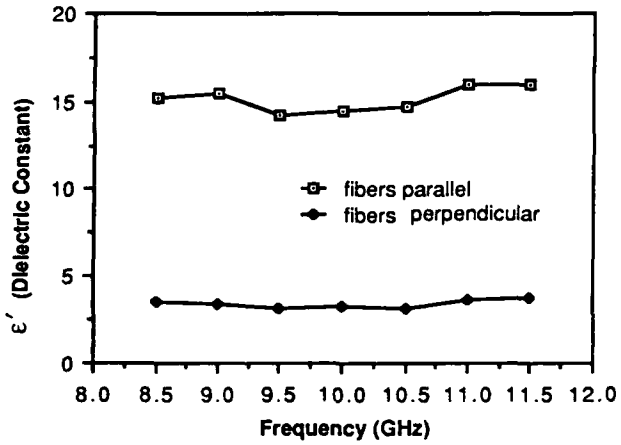


Fig. 4 - Dielectric Constant vs Frequency for Ni-coated C Fibers in Epoxy. Measurements with Fibers Parallel and Perpendicular to the Electric Field (Aspect Ratio 80-1).

For these fiber samples the orientation distribution was carefully analyzed optically and is displayed in Fig.5. Because of the flow conditions during molding, any variation in fiber orientation occurs primarily in the plane parallel to the thin direction. This is the plane to which the electric field is confined in the dielectric measurements. The extent of variation out of the plane is indicated in Table 2. Also included is the fraction of fiber to fiber contact observed and the resistivity both parallel and perpendicular to the preferred fiber direction. As expected, the resistivity is high because of the small volume percent fiber but, is lower parallel to the fiber rather than perpendicular. This seems to be consistent with the very small amount of fiber to fiber contact observed.

To determine the applicability of equations 6 and 7 to the measured data one must consider the orientation distribution shown in Fig.5 and, to a first approximation, assume no fiber to fiber interaction. Under these conditions the effective dielectric constant both parallel and perpendicular to the preferred fiber direction is given by

$$\epsilon_1^*, \epsilon_2^* = \frac{1}{n}\sum_{i=1}^{n}\left[\epsilon_1 cos^2\theta_i + \epsilon_2 sin^2\theta_i\right] \qquad (9)$$

where ϵ_1 and ϵ_2 are the dielectric constants given in equations 6 and 7 respectively, ϵ_1^* and ϵ_2^* are the two effective dielectric constants for the fiber distributions measured, θ_i is the angle between the field and fiber direction and n is the total number of fibers in the distribution studied.

Table 2
Data for Oriented Fiber Samples
Aspect Ratio = 80

	Sample of Fig. 5a	Sample of Fig. 5b
Fibers out of plane < 10°	< 3%	< 6%
Extent of Fiber to Fiber Contact	< 5%	< 9%
Resistivity-ohm cm		
Parallel to Fiber	1.4×10^{12}	2.8×10^{11}
Perpendicular to Fiber	4×10^{13}	3×10^{13}

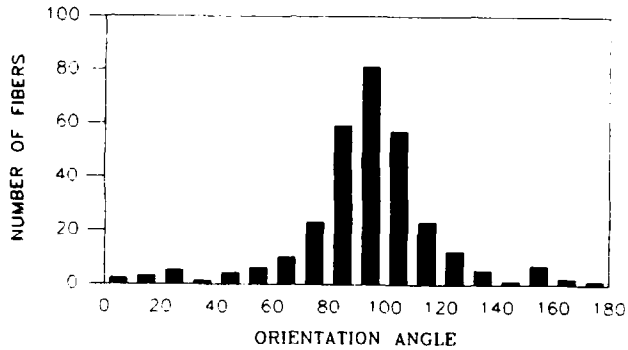


Fig. 5a. - Fiber Orientation Distribution for Sample with Fibers Parallel to Electric Field. Field at Angle 90°.

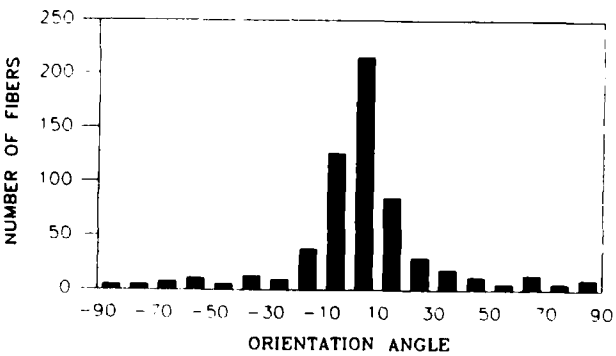


Fig. 5b. - Fiber Orientation Distribution for Sample with Fibers Perpendicular to Electric Field. Field at Angle 90°.

From the measured dielectric data for the two directions and the orientation distribution of Fig.5, the value of ϵ_1 and ϵ_2 are found to be 17.39 and 1.19 respectively. These can be compared with the values of ϵ calculated from equations 6 and 7 for an aspect ratio of 80. For this aspect ratio h(a) is calculated to be 0.99936 leading to dielectric constant of 13.96 and 2.82 [2]. Some error enters because of the need for an extremely accurate value of h when the aspect ratio is large and $h \rightarrow 1$. The fit is reasonable giving validity to the approach leading to equations 6 and 7. A better evaluation would be possible if fibers of smaller aspect ratio could be used and oriented. However this poses more difficult problems in composite synthesis.

For ferromagnetic particles it is possible to use magnetic fields to affect some degree of alignment of fibers and also particulates. This has been done for both nickel coated fibers in epoxy as well as nickel particulates in both epoxy and wax. With the particulates anisotropic behavior occurs as the particles tend to align in the magnetic field. Some preliminary results for the particulates are shown in Table 3. An example of the extent of alignment can be seen in the microstructure of Fig.6. The effective aspect ratio resulting from the magnetic field is significant. To date, sufficient data has not been collected to explore the magnetic field effect thoroughly.

Table 3
Dielectric Data (8.5 - 11.5 GHz) for Random and Oriented Ni Particulates (2.2 - 3μm) in Wax

Vol.%	Dielectric Constant		
	Random	Aligned	
		Parallel To Field	Perpendicular To Field
2.9		11.82	
3.1			8.85
3.4	6.02		

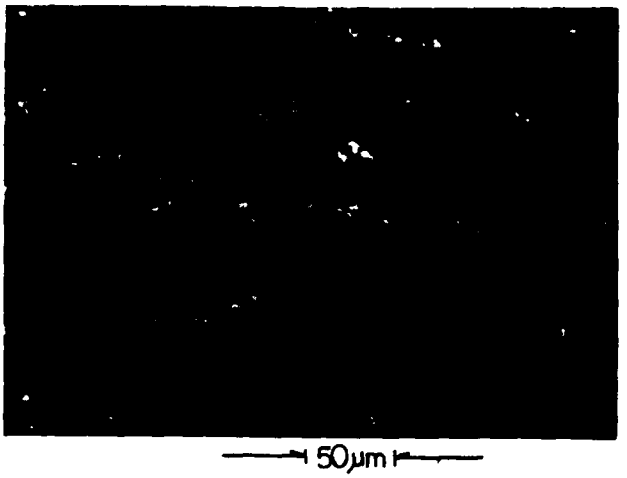


Fig. 6 - Microstructure of Ni Particulates (2.2 - 3.0/μm) in Epoxy. Aligned in Magnetic Field (.78 Vol %).

D.C. conductivity and microwave dielectric measurements were also made on randomly oriented fibers in polyester. These results are summarized in Table 4. Even though the volume fraction of fiber is low (0.002) the dielectric constant is large, much greater than for the oriented fiber sample of approximately the same volume fraction. This is due primarily to the higher aspect ratio of the fibers, but possibly also an appreciable amount of fiber to fiber contact which is evident in the high D.C. conductivity. The high conductivity is also reflected in the high imaginary part of the dielectric constant and tan δ measured for these samples.

Table 4
Data for Random Oriented Fiber (.002 Vol %)
Aspect Ratio = 800

Sample	Dielectric Constant		Resistivity
	Real	Imaginary	ohm cm
1	33.23	10.6	170
2	25.1	11.0	170

A measurement of the dielectric properties of these random samples was attempted at 10KHz with a capacitance bridge. The sample appeared to act only as a conductor. This is expected when the conductivity is high. Since the imaginary part of the impedance increases with frequency the complex capacitance of such samples are more easily measured at higher frequencies.

A simple attempt to analyzethe dielectric results for these random fiber samples was not successful. If equation 6 and the value of Y for an aspect ratio of 800 (.99999) is used a dielectric constant of 557 is obtained as a possible upper limit. This contrasts with the measured values of approximately 30. No simple orientation averaging would reduce the calculated value to this low number. It is expected that fiber to fiber contact and fiber interations will influence the measured results, but also equation 6 is very sensitive to y values close to one.

SUMMARY AND CONCLUSIONS

Microwave dielectric and D.C. conductivity measurements were performed on a series of metal-filled nonconducting matrix composites of different metal type, particle size and matrice. For the metal particulate samples the volume fraction of metal extended up to and beyond the expected value for percolation. Studies on metal fiber composites were confined to a low volume fraction but included a consideration of fiber aspect ratio and orientation. This study led to the following observations and conclusions.:

1. No simple mean field or effective medium theory can explain the variation of the microwave of dielectric constant with volume fraction of particulate.

2. The dependence of dielectric constant on particulate size strongly suggests the possibility that interfacial electical effects at the metal-matrix boundary are important considerations.

3. The results for different metal particulates and different matrices seem to follow the same trend.

4. No evidence of percolation was found even for particulate volume fractions in excess of the expected percolation value. This suggests the possibility of an electrical barrier at the point of particle to particle contact for all of the systems studied.

5. For the oriented fiber samples the dielectric results at low volume fraction are consistent with an approach that takes into account both the fiber orientation and its aspect ratio.

6. Randomly oriented fiber samples show a higher tan δ value which is consistent with a higher D.C. conductivity associated with appreciable fiber to fiber contact.

ACKNOWLEDGEMENT

This work was supported by the office of Naval Research under Contract NO N00014-86-K-0026 and the Center for Composite Materials, University of Delaware.

REFERENCES

[1] J.C. Maxwell, "Electricity and Magnetism", Oxford University Press, London (1892)
[2] See for example reviews by D.K. Hale, J. Mat. Sci. 11, 2105(1976); G. Banhegyi, Colloid and Polymer Sci. 264, 1030 (1986); R.L. McCullough, Composite Sci. and Tech. 22, 3 (1985)
[3] J.C.M. Garnett, Philos, Trans. R. Soc. 203, 385(1904); 205, 237(1906)
[4] D.A.G. Bruggeman, Ann. Phys. Leipzig 24, 636 (1935)
[5] H. Looyenga, Physica 31, 401 (1965)
[6] R.W. Cohen, G.D. Cody, M.D. Coutts, B. Abeles, Phy. Rev. B8, 3689 (1973)
[7] for a review see D. Stauffer, Phys. Rep. 54, 1 (1979)
[8] I. Webman, J. Jortner, M.H. Cohen, Phys. Rev. B12, 5712 (1976)
[9] R.B. Laibowitz, Y. Gefen, Phys. Rev. Lett. 53, 380 (1984)
[10] I. Chen, W.B. Johnson, J. Mat. Sci. 2[illegible] (1986)
[11] Y. Song, T.W. Noh, S. Lee, J.R. Gain[illegible]v. B33, (1986)
[12] C.T.D. Wu, R.L. McCullough, "D[illegible]oments in Composite Materials", edited by G. [illegible]oister, Applied Science Publishers (1979)
[13] M.H. Cohen, Geophysics 46, 1057 (1981)
[14] J.R. Macdonald, Phys. Rev. 92, 4 (1953)
[15] C. Wagner, J. Phys. Chem. Solids 33, 1051 (1972)
[16] P. Sheng, E.K. Sichel, J. I. Gittleman, Phys. Rev. Lett. 40, 1197 (1978)
[17] P. Sheng, J. Klafter, Phys. Rev. B27, (1983)
[18] S.H. Kwan, F.G. Shin, W.L. Tsui, Mat. Sci. 15, 2978 (1980)
[19] Y.S. Ho, J.J. Kramer, Mat. Res. Soc., 1988 Spring Meeting, April 5-9, Reno Nevada (in press)
[20] Alain Priou, private communication
[21] K.T. Chung, A. Sabo, A.P. Pica, J. Appl, Phys. 53, 6867 (1982)
[22] Y. Gefen, W. Shih, R.B. Laibowitz, J.M. Viggiano, Phys. Rev. Lett. 57, 3097 (1986)
[23] R.L. McCullough, private communication

MATERIALS AND PROCESSES OF FLEXIBLE PRINTED CIRCUITRY

O.C.C. Lin, J. M. Liu
Materials Research Laboratories, ITRI
Hsinchu, Taiwan 31015, R.O.C.

ABSTRACT

Flexible printed circuitry has been widely used since the early 1970s in military applications as a critical component of high reliability electronic device. During the past decade, its application has spread rapidly to personal computers, computer peripherals, and home and consumer electronics. This remarkable achievement can be attributed to the unique features that flexible printed circuit provides: three-dimensional space utilization, formed and shaped circuits, movable circuit components, and control-connector integrations. Polymer materials designed as circuit substrates providing these features are mainly polyimides and polyesters. Of key importance to the material system is polymeric adhesive which bonds the flexible substrate and the conductor lines at all times during the processing and end-use performance of the circuitry. Major processing steps of the circuitry includes lithographic developing and stripping, high speed drilling, plasma etching, high temperature soldering and lamination under high pressure and elevated temperature. These processes encompass intricate interplays of the physical, chemical and rheological properties of the materials under various external conditions. Acrylic copolymers are known to have the best balance of properties. Many important parameters of flexible printed circuit materials and processing will be discussed.

THE APPLICATION OF FLEXIBLE PRINTED CIRCUIT

(FPC) started with high precision, high reliability military equipment in the U.S.A. during early 1970s. Its contribution to the surface-to-air missiles and other weapon systems were well known.(1) Military equipment remained to be the dominant user of FPC until 1980s. At that time, it accounted for 60% of the high quality FPC used. The remaining 40% was splitted among computer and various other end uses. Whereas military applications grew steadily over the last decade, industrial applications expanded with a more rapid pace. It grew from main frame and personal computers to computer peripherals such as printers, disk drivers and to consumer electronics such as camera and video cassette recorders. At present, over 60% of the end uses are in the industrial and consumer electronics sectors. The FPC market forecast in United States will reach over US$500 millions in 1988 representing an average annual growth rate of 14% over a 5-year period. The growth rate is higher in Japan due to the faster growth in consumer electronics utilization (Table 1).

The remarkable growth of FPC can be attributed to the unique characteristics that FPC provides:

- three dimensional space utilization
- formed and shaped circuits
- flexible and movable circuits parts
- integrated control and connector

Thus, by the use of flexible printed circuit, the total value in-use of a product is much improved. The "miniaturization" of many electronic products is made possible by the use of

FPC. A good case worth mentioning is the modern automatic pocket camera. (2) All the focusing, lighting, film advancing and rewinding are controlled by many microprocessors. These microprocessors are attached to a piece of flexible printed circuit wrapped , twisted, and fitted into the very small opening boundaried by the various mechanical parts. Some typical end uses of FPC are shown in Table 2.

FLEXIBLE PRINTED CIRCUITRY MATERIALS

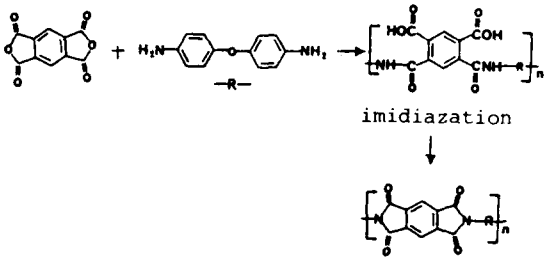
THE BASIC STRUCTURE (Fig.1)- The basic components that consititute a FPC are:

- copper clad - copper foil laminated to flexible substrate by high quality adhesive. The copper foil can be laminated to one or both sides of the substrate forming a single-sided or double-sided clad. Electronic circuits are to be etched out of copper foil.
- coversheet - a polymer film with adhesive attached. The coversheet is laminated to the circuit for mechanical and electrical protections. The polymer film and adhesive are chemically identical to that in the copper clad with the exception that the adhesive is in the "B-staged" form, or partially reacted form, whereas the adhesive in the copper clad is at the "C-staged" form, or fully reacted, or cured form.
- sheet adhesive - This is the stand-alone "B-staged" adhesive, identical to that in the coversheet, used to bond several etched copper clads together forming multilayer printed circuit boards.

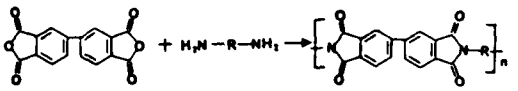
These components are similar to those of rigid printed circuit board. The major exception is the substrate . With rigid printed circuit(RPC), the substrate is usually glass fabric impregnated with polymer resin. The RPC sheet adhesive, called bond-ply, is made of the "B-staged" resin and glass fabric composite. Table 3 compares the various components of the flexible and rigid printed circuits. The basic components: polymer film, polymer adhesive and copper foil will be further discussed below.

THE POLYMER SUBSTRATE - Polymer materials designed as flexible circuit substrates providing the necessary features are mainly polyimides and polyesters. Polyimide-based FPC dominates the market both in dollars and in volumes. Polyimide is the most successful high temperature polymer commercially available. Its film is the only known film so far which can maintain its integrity after soldering at 260°C. The "Kapton" film marketed by the Du Pont Company has long been the leader of its kind. However within the past five years, several companies, for example, Ube Industrial Corporation has entered into this area employing varying chemical modifications. The difference in chemistry among them is illustrated as follows.

"Kapton"-like



Upilex



It should be noted that "Kapton" -like polyimide film is more familiar to the end users and is thus widely adopted.

THE ADHESIVE - Adhesive is the only material which undergoes changes in property during the FPC fabricating processes. It plays a pivotal role in FPC technology. The important parameters for a good FPC adhesive are: (a) good adhesion to the copper foil and the polyimide film, (b) dielectric properties, (c) solder resistance, (d) chemical resistance, (e) low flow property, (f) flexibility, (g) dimensional stability, (h) curing temperature, (i) storage lifetime. A delicate balance must be maintained among these parameters. An illustration on how a good adhesive can be obtained through optimizing molecular design is given below.

Dimensional stability is very important for fine line and large

circuits. For unreinforced FPC material it is important to have a low shrinkage adhesive before and after cure. These requirements imply a formulation with low degree of crosslinking and that is consistent with the need of flexibility. A flexible, low flow and dimensionally stable adhesive indicates a linear high molecular weight polymer with small crosslinkable functional groups. However, it is difficult to have a solvent soluble polymer which exhibits excellent solvent resistance with only slight degree of crosslinking. One way to solve this dilemma is to employ polymer latices having irreversible film-forming mechanism. Linear high molecular weight polymers synthesized by emulsion polymerization is not soluble but dispersible in water, and is solvent-resistant after film formation. The solvent resistance may be further improved by slight corsslinking reactions during lamination process.

Solder resistance , another important parameter, correlates to the thermal stability of the cured adhesive which, in turn, is dependent on the molecular structure and high molecular weight of the polymer. And for better adhesion, attaching suitable functional group on the polymer will help. Experimentally, better flow property and low shrinkage after cure will result in better adhesion. So in the molecular design of the adhesive, the curing temperature and the flow property have to be matched carefully.

Considering these intricate property requirements, it is little wonder that only epoxy and acrylic polymers have been successfully used in FPC applications. Epoxy adhesives are widely employed in Rigid Circuit Board. However, its flexibility and bonding strength is not sufficient for certain high end applications. Acrylic polymer synthesis consisting of multicomponents with complementary functions has been demonstrated to be the better choice. The Pyralux series of Du Pont Company is a well known examples.(3)

Recently, a single component emulsified acrylic polymer has been developed as adhesive for FPC.(4) The polymer is designed to have hard and a soft segments with a functional and a crosslinkable group. Examples of the constitutive groups are:

soft component- butylacrylate,
2-ethylhexylacrylate
hard component- methylmethacrylate,
acrylonitrile, styrene
functional component- 2-hydroxyethyl
acrylate, acrylamide
crosslink component- glycidyl
acrylate

By manipulating the soft and the hard component, a thermally stable, solvent resistant polymer with excellent film forming property can be synthesized. Likewise, by changing the functional and the crosslink components a high conversion, low coagulation and stable latex adhesive with matched flow and curing properties can be made. It was also found that batch type and controlled-feed semicontinuous emulsion polymerization methods (5) resulted in adhesives with different rheological properties (Fig 2). Further polymer with random distribution of monomer components has better flow property before cure and better mechanical strength after cure.

The apparent combination of methacrylic acid and glycidyl methacrylate as functional component and crosslink component proved to be inadequate. The acid group and the glycidyl group tend to react under the emulsion polymerization conditions with emulsifier (surfactant) acts as catalyst.

The adhesive development work had been carried on further to make rolled copper clad. It is found that the molecular design concept may be tailored to meet the need of an adhesive for continuous roll-to-roll lamination. The properties of the adhesive and the copper clad laminates are summarized in Table 4.

COPPER FOIL - Electrodeposited copper foil (ED copper) is commonly used in rigid printed circuitry . While one side of the foil has a very smooth surface, the other side is relatively rough. Physical and chemical treatments are needed to enhance adhesion. ED copper has cellular grain structures perpendicular to circuit direction, which can cause inadequate elongation and ductility for fast moving components such as magnetic head.

Rolled annealed copper (RA copper) on the other hand, is made by casting and rolling processes of high purity copper. Its structure is uniform and its crystal direction parallels to the circuit rendering high ductility and

elongation properties. RA copper is suitable for FPC utilization. Its price is usually higher. The choice of ED copper or RA copper should depend on the end-use application. ED copper with high ductility have also been on the market for sometime too. They are rigid at ambient low temperature thus easy to handle, and flexible after heat treatment at 180 They are gaining more utilization in the FPC industry.(6) A comparision of ED copper, RA copper and highly ductile ED copper are show in Table 5.

FLEXIBLE PRINTED CIRCUITRY PROCESSING

The processing scheme of FPC is generally similar to that of RPC. A typical processing flow chart for a 4-layer flexible printed circuit is shown in Fig.3. Due to the different properties in materials employed, special precautions are required for FPC. Some highlights are described as follows:

STRIPPING/ETCHING - Owing to the nature of FPC, circuits are subject to distortions due to the chemical degrading or solvent swelling of adhesive and/or polyimide film. Among those processing chemicals for FPC manufacturing, stripping chemicals deserve specical attention. It is found that prolonged contact with organic strippers can cause varying degrees of film distortion, blistering and film separation, etc. The distortion can be minimized by lowering spray pressure and by reducing residence time at the stripping bath. Thorough rinsing followed by complete drying is of utmost importance for the removal of adsorbed/absorbed stripper which may cause blistering and loss of adhesion with repeated lamination. For optimum stripping, the following systems are recommended.(7)

-CH_2Cl_2/CH_3OH
  by dipping, 2 min. 80°F
  by spraying, 1 min. 80°F

-"Riston" S-1100X
  by dipping, 2 min. 130°F
  by spraying, 2 min. 130°F
-NaOH,2%
  by dipping, 6 min. 120°F

PLASMA DESMEAR OF PLATED-THROUGH-HOLE (PTH) - The quality of PTH in a printed circuit board determines the effectiveness of electrical interconnections, and consequently, the success or failure of the circuit. A good PTH is the integrated result of three individual processes: hole drilling, hole cleaning, and electroless/ electrolytic plating.

Although the use of aluminum foil entry materials and aluminum clad backer material was beneficial for retardation of burr formation, it was found that no drilling conditions could prevent drill smears. There are two methods commonly used for desmearing. The wet method involes the use of a combination of acids which oxidizes and dissolves polymeric residues, i.e. drill smears on the wall of the holes. Of all the acids commonly used, chromic acid is the most effective in desmearing FPC with acrylic adhesives. However, chromic acid can be absorbed by the adhesive causing swelling and other undesirable effects for the subsequent plating process. Further, chromic acid is toxic and presents difficult handling and disposal problems. Thus, the wet method has quickly given way to the dry method, the desmear and etchback of drilled holes by plasma.

Plasma desmear and etchback of drilled holes of multilayer FPC has been shown to be successful using both parallel and barrel type equipments. The effects of plasma etching with a mixture of CF_4/O_2(30/70) on the PTH are in Fig 4.

The mechanism of plasma ethching in FPC is attributed to physical attrition and chemical reactions. The details of the latter are very complicated. It is suggested the etching process starts with the generation of fluorine and oxygen free radicals by RF glow discharge. The fluorine free radical is very active and decomposes polymer residues to low molecular weight intermediates or monomers. The decomposed products are further oxidized by oxygen free radicals to gaseous compounds which are then evacuated.(8)

It was known that strong alkaline solvent had a tendency to swell the acrylic adhesive resulting in plating problems. Although no attack of the acrylic materials was observed with acid cleaners, marginal adhesion to the polyimide dielectric materials was

noted. By using the conventional alkaline cleaner at a lower temperature (125 for a shorter period of time (2 minutes), excellent hole wall adhesion was consistently obtained.(9)

COVERSHEET LAMINATION - Coversheet lamination which entails encapsulating the etched flex circuit with diclectric materials is unique to FPC. Improper drilling and coversheet lamination can lead to wicking of solder under the coversheet and around access pads drilled or punched during the solder operation. This delamination or non-lamination of the coversheet may lead to chemical corrosion of the circuit by solder flux or even shorting of the signal between adjustment pads.

For circuits using one or two mil polyimide film in the coversheet, solder wicking is readily eliminated by increasing the lamination pressure or changing the components of the press pad systems. For circuits using three or five mil polyimide film in the coversheet, the problem is more complicated. A Plackett- Burman screening designed for evaluating the various parameters was carried out. Testing and data analysis have shown that solder wicking can be avoided by (a) eliminating polyimide film distortion and adhesive tear-out caused by poor drilling or punching conditions, (b) use of two mil adhesive instead of one mil, (c) changing press pad system and, (d) proper copper preparation of etch copper details. (10)

FUTURE TRENDS

At present time, batch lamination process are used to make copper clads for high performance application,i.e. military and aerospace utilization. In non-military area, rolled copper clad is preferred by end-users because it offers versatility in circuit design and material utilization efficiency. However, manufacturing rolled clad with good dimensional stability and high peel strength comparable to those of sheet copper clads remains to be a challenge.

The successful utilization of FPC in consumer electronics results in large volume circuit fabrication. Thus, roll-to-roll continuous processing of FPC is more attractive. By far continuous fabrication of FPC is associated to photolithographic process only. With continual efforts in research and development, a fully continuous and automatic production line from rolled clad to circuitry and to assmebly of electronic components will become standard sometime in the future.

TAB (Tape Automated Bonding) is one of the most promising high density IC packaging technologies that can be used in conjunction with surface mount technology. An IC component of TAB packaging is manufactured by (a) punching PI film to form specific pattern, (b) rolled lamination of copper foil to the preshaped polyimide film, (c) photolithographic process to form specific pattern on copper foil, (d) chip mounting and wire bonding. The roll-to-roll lamination and processing techniques of FPC can readily be applied to TAB fabrication.

The application of multilayer rigid-flex circuit are gaining popularity. Rigid-flex is a hybrid of RPC and FPC in one PC board. In rigid-flex, the main body is usually a multilayer rigid board with some conductor layers made of single or double-sided FPC. Three dimensional space utilization can be achieved by bending or twisting of the flexible part. Reliability and high density packaging is thus achieved at one time. Rigid-flex was limited to military utilization because of high cost. Recently, with the cost reduction in FPC, rigid-flex begins to debut in the industrial application of computers and telecommunication systems. The application of rigid-flex circuits will become increasingly popular in the future.

ACKNOWLEDGEMENT

This work was supported by a grant from the Ministry of Economic Affair, the Republic of China under contract No. ITRI-047-P301 to ITRI.

REFERENCES

1. The Du Pont Magazine, P.4, November/December, 1981
2. Kachihiko Kōtō, Electronic Packaging Technology (Japanese) 1, 144-146(1985).
3. Yuan E., U.S. Patent 3,822,175 (1974)
4. Liu, J. M., F. L. Chen and Y. C. Chiou, Chinese Patent pending; U. S. Patent pending
5. (a) K. Chujo, Y. Harada, S. Tokuhara, and K. Tanaka, J.

Polym. Sci. Part C, 27,321(1969)
(b) S.C. Misra, C. Pichot, M.S. El-Aasser, and J.W. Vanderhoff, J. Polym. Sci. Polym. Lett. Ed., 17,567(1979).
(c) S.C. Misra, C. Pichot, M.S. El-Aasser, and J.W. Vanderhoff, J. Polym. Sci. Polym. Chem. Ed., 21,2383(1983).
6. Kawada, Junichiro, Function & Materials 4, 37-44(1984)
7. Lin, O. C. C. unpublished results.
8. Lin, O. C. C. and E. Yuan, Proc. of Inst. Interconnecting & Packaging Elec. Circuit, Fall Meeting, Dallas, 1981, and IPC-TP-369.
9. Hayes, L. E. " Manufacturing Multilayer Rigid-Flex Hardnesses", Report DAAK 40-78-C-0244(1981).
10. Robertson, S. S., O. C. C. Lin, E. L. Yuan, H. E. kime, J. C. Chevrier, and W. J. Lautenberger Proc. of Inst. Interconnecting & Packaging Elec. Circuit, Fall Meeting, Hollywood, 1983, and IPC-TP-457.

Table 1. FPC Market and Growth Rate

Year / Country	1983	1987	1988	Average Annual Growth Rate
USA 1 million USD	265 [a]	—	504 [a]	14%
Japan 100 million ¥	303.9 [b]	612.3 [b]	786 [c]	21%

a IPC/IMRC, 1984
b Nikkei Electronics, April, 1986
c Japan Printed Circuit Association, 1987

Table 2. Advantages and Application of FPC

Advantage	Application Examples
Three dimensional space utilization	camera, cassette tape recorder, missile, transceiver
Flexibility	printer, hard disk drive, floppy disk drive, recorder, CD player
Shaped and formed circuit	facsimile(FAX), medical equipment, telephone
Integrated connector & circuits	car dashboard, harness and cables for missiles

Table 3. Comparison of RPC and FPC Materials

	Rigid PC Board	Flexible PC Board
Substrate	Phenolic/paper Epoxy/glass Polyimide/glass Metal core	Polyimide Polyester
Adhesive	Epoxy Phenolic Polyimide	Acrylic Epoxy
Conductor	ED copper High ductility copper	ED copper High ductility copper RA copper

Table 4. Properties of MRL Flexible Printed Copper Clad

	Sheet Copper Clad	Rolled Copper Clad	Test Method
Peel Strength, 90° (lb/in)	11-13	10-13	IPC-650-2.4.9.
Dielectric Constant (1MHZ)	3.7	3.7	ASTM-D150
Dissipation Factor (1MHZ)	0.026	0.026	ASTM-D150
Dielectric Strength (KV/mil)	3.3	3.3	ASTM-D149
Volume Resistivity (Ω-cm)	4.7×10^{15}	4.7×10^{15}	ASTM-D257
Surface Resistivity (Ω)	4×10^{15}	4×10^{15}	ASTM-D257
Chemical Resistance	no change	no change	IPC-650-2.3.2.
Dimensional Stability (mil/in)	0.2-0.6	0.4-1.3	IPC-650-2.2.4.
Solder Float, 260℃, 30 sec.	no blistering no delamination	no blistering no delamination	—

Table 5. Properties of Rolled and Electrodeposited (ED) Copper

Copper Foil	Condition	Copper Weight	Tensile Strength (psi)	Elongation % in 2 Inch
Rolled-K	Annealed	1 oz.	30,500	20
Rolled-bare	Annealed	1 oz.	28,500	20
ED	Standard	1 oz.	50,000	3
ED	Annealed	1 oz.	20,000	10
ED	High temp. elongation	1 oz.	50,000	15

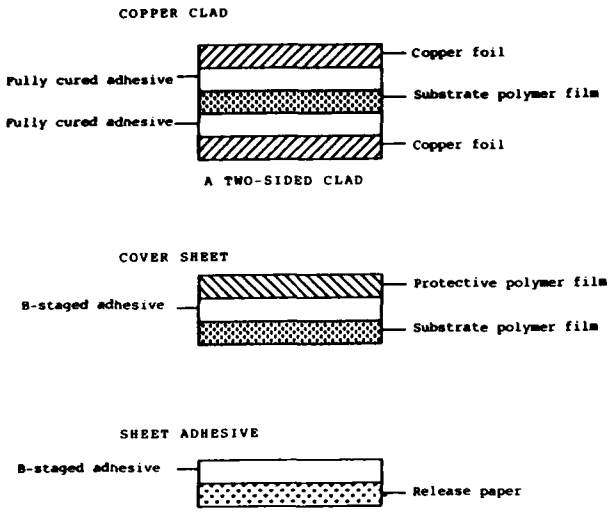


Fig.1 Structure of FPC Materials

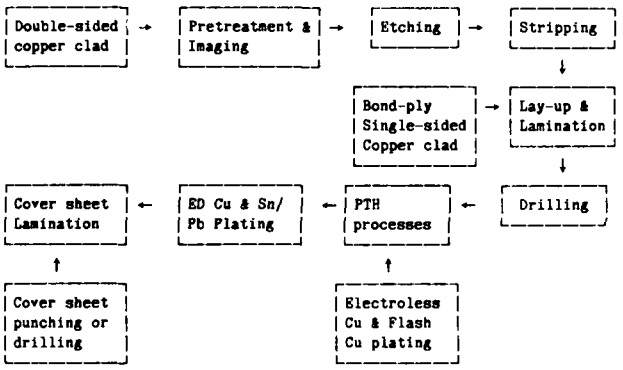


Fig.3 4-Layer FPC Processes

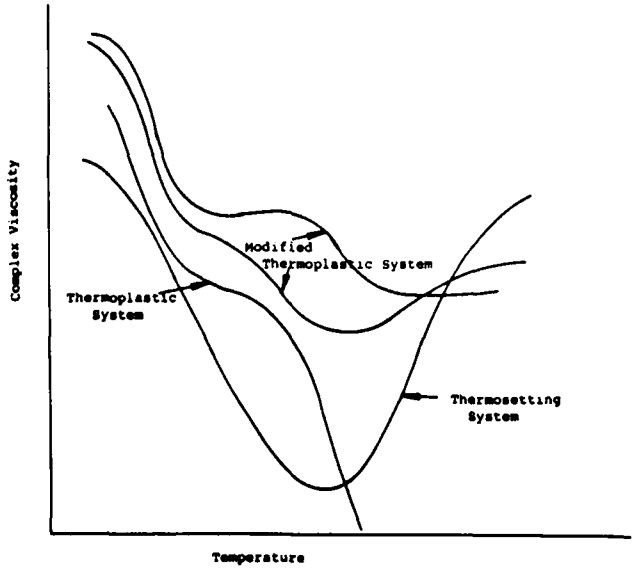
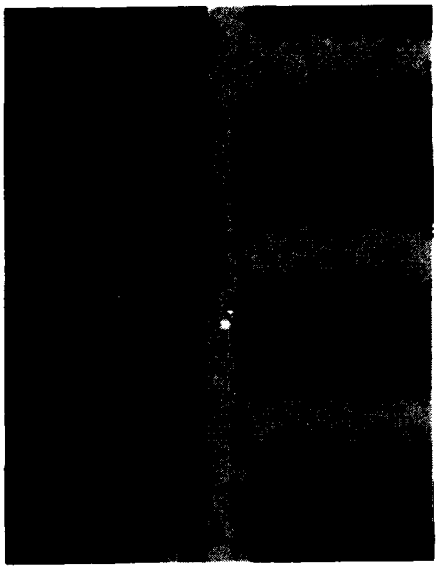
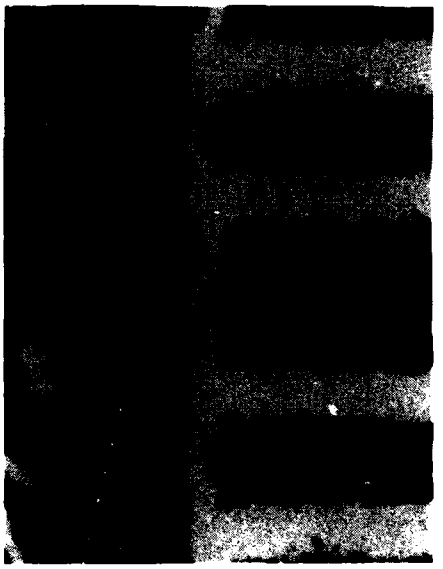


Fig.2 Schematic diagram showing different types of rheological behaviors of polymer systems designed as FPC adhesive.



(a)



(b)

Fig.4 PTH micro-section, (a) unetched (b) plasma etched

HYDROGENATED AMORPHOUS SILICON: A NEW MATERIAL FOR LARGE AREA ELECTRONICS

R. A. Street
Xerox Palo Alto Research Center
Palo Alto, California 94304, USA

ABSTRACT

Flat panel displays, solar cells, printers, and photocopiers, all require electronic circuits that extend over a much larger area than is possible with crystalline semiconductors. Amorphous semiconductors have gone a long way to meeting this need, particularly hydrogenated amorphous silicon (a-Si:H). As its material properties have become better controlled, the market for a-Si:H devices has expanded rapidly. Hydrogen is a crucial component of a-Si:H, as it reacts with broken or strained silicon bonds to reduce the density of electronic defects. *Amorphous silicon* can be substitutionally doped, so that junction devices can be made, although the physics of the doping is qualitatively different from that of crystalline silicon. This paper discusses material properties and device fabrication issues.

HYDROGENATED AMORPHOUS SILICON has now been studied for about 20 years. Prior to that time, pure amorphous silicon was found to have such a large defect density that it was unusable for electronic devices. The beneficial effects of hydrogen were discovered at least in part by accident, when the material was deposited from silane (SiH_4) gas in a plasma discharge.[1] This method of growth results in a reduction of the defect density by four orders of magnitude, giving material that is of device quality. It is now clearly understood that hydrogen removes defects by bonding to unterminated silicon atoms. The first demonstration of substitutional doping, made possible by the low defect density, was reported in 1975 and opened the way to many device applications.[2] Since that time the research effort has greatly expanded so that this material now dominates studies of amorphous semiconductors, and many applications have developed. The first solar cells were reported in 1976,[3] and the conversion efficiency has steadily increased to its present value of 13-14%. In 1979 the plasma deposition of silicon nitride was used in conjunction with a-Si:H to produce field effect transistors.[4] Large area arrays of these thin film transistors (TFT) are now important in liquid crystal arrays,[5] and monolithic circuits for printing and input scanning applications.[6] The combination of doping and transistor action, means that essentially all of the circuit elements used in crystalline silicon electronics can be reproduced in a-Si:H, giving it a broad versatility in electronic circuit design. The principal advantage of using a-Si:H is that it can be deposited over large areas on low cost substrates such as glass. A disadvantage of a-Si:H is that its low mobility sometimes leads to a slow performance. The applications are therefore primarily aimed at electronic devices in which large size is essential, of which liquid crystal displays, printers, scanners, detector arrays, and solar cells are obvious examples.

GROWTH AND MATERIAL PROPERTIES

Fig. 1 shows a schematic diagram of the most common configuration for the growth of a-Si:H.[7] The plasma chamber is a capacitor structure driven by a radio frequency plasma. Silane (SiH_4) is introduced at low pressure and dissociates in the plasma. The reactive gas radicals - SiH_3, SiH_2 etc.- are adsorbed on the surface of the growing film resulting in the a-Si:H layer. Good quality films require a fairly careful choice of plasma conditions (eg temperature, pressure, r.f. power etc.).[7] For example a sufficient surface mobility of the atoms is needed to ensure that all the covalent bonds are satisfied to give a low defect material. Since the best material contains about 10 at. % hydrogen, most of the hydrogen that arrives at the surface, in the form of SiH_3 etc. is reemitted into the gas. If the temperature of growth is too

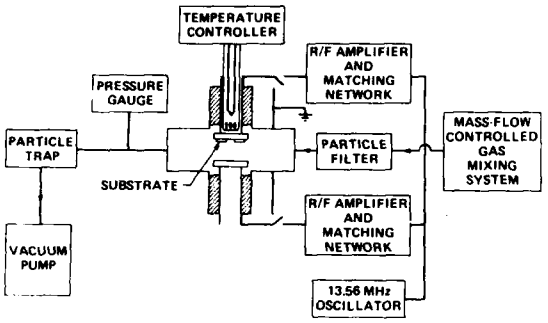


Fig.1. Schematic diagram of r.f. plasma apparatus for the deposition of a-Si:H.

low, the surface mobility is insufficient for the hydrogen to properly terminate the defects, but a temperature too high results in too much of the hydrogen being lost. Generally the best films are grown at 200-300°C.

The electronic properties of any semiconductor are governed by the conducting states near the band edge, and by the localized states in the gap. The density of states distribution is illustrated in Figs. 2 and 3.[8] In amorphous semiconductors, the absence of a periodic crystalline structure results in a high density of localized states very close to the band edges. The density of these band tail states decreases exponentially into the gap. The states that result from broken bonds or other specific defects are deeper in the band gap.

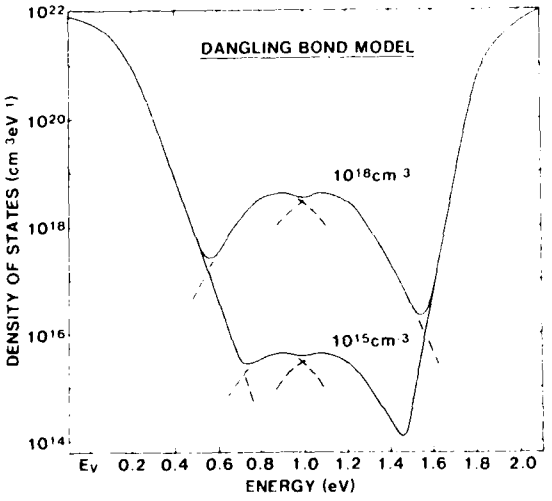


Fig. 2. Density of states model showing the band tail and deep defects. The two curves correspond to different total densities of defects.

The band tail states control the electrical conduction. Carriers are constantly being trapped and released by the localized states, resulting in an electron mobility of only 1 $cm^2/Vsec$, and a hole mobility even lower. In contrast, crystalline silicon has free electrons in the conduction band with an electron mobility of about 1000 $cm^2/Vsec$. Since the band tail originates from the disorder inherent in the amorphous network, the low mobility appears to be a fundamental limitation of amorphous semiconductors, although some improvement may be possible by reducing the bonding disorder.

Electrical transport occurs at a mobility edge denoted by the energy E_C, which is near the edge of the band tail as indicated in Fig. 3. Even in doped material the Fermi energy, E_F, remains within the localized states so that the conductivity, σ, is thermally activated,[9]

$$\sigma = \sigma_o \exp[-(E_F-E_C)/kT]$$

where σ_o is the conductivity at the mobility edge, and has a value of about 100 $\Omega^{-1}cm^{-1}$. Fig. 4 illustrates the temperature dependence of the conductivity in n-type and p-type a-Si:H, and shows that σ is much lower than in the crystal.[10] Despite many years of work, the details of the conduction near the mobility edge are still not well understood and remain an intriguing problem in solid state physics.

The defect states lie deep in the band gap with a density that depends on the growth conditions, as indicated in Fig. 2. These states determine the recombination of excess carriers and so influence the photoconductivity. The dominant defect in undoped a-Si:H is the neutral silicon dangling bond.[11] Over recent years, various spectroscopic measurements have accumulated much information about the properties of dangling bonds. For example the defects are observed by electron spin resonance (ESR) to have a characteristic g-value of 2.0055.

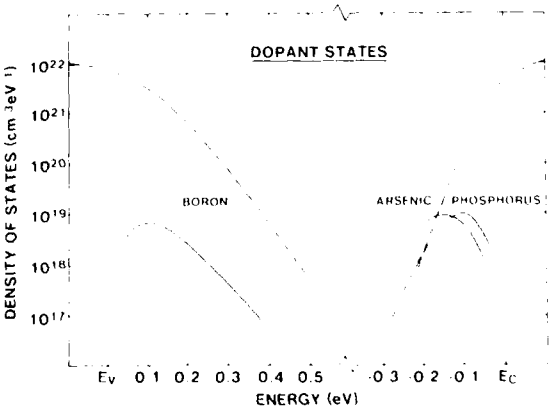


Fig. 3. Density of states distribution showing the band tail region and the broad donor and acceptor bands. The densities are shown for a doping level of 10-3.

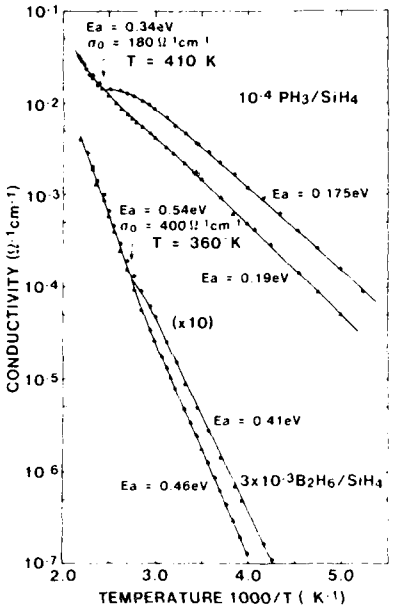


Fig. 4. Temperature dependence of the d.c. conductivity of n-type and p-type a-Si:H, after annealing and cooling at different rates. Faster cooling results in a higher conductivity. The thermal equilibration temperatures are indicated.

The associated levels in the gap act as deep traps and recombination centers and are observed in luminescence, optical absorption, and transient photoconductivity. In undoped a-Si:H, the defect density can be reduced to about 10^{15} cm^{-3} by a suitable choice of preparation conditions.

SUBSTITUTIONAL DOPING

Substitutional doping is an essential property for semiconductor devices. Dopants are introduced into a-Si:H by adding PH_3 or B_2H_6 into the deposition gas usually in concentrations ranging from 10-6 to 10-2. N-type doping is more effective than p-type in increasing the conductivity, with a change at room temperature of a factor 10^6 compared to undoped a-Si:H.[2] The binding energy of donors and acceptors are larger than in the crystal, and the energy levels are broadened by the disorder. The location of the donor and acceptor bands shown in Fig. 3, indicates that arsenic is a deeper donor than phosphorus.[8] Although doping occurs in a-Si:H, the underlying doping mechanism is qualitatively different from that in crystalline silicon.[12] One major difference is that the dangling bond density increases rapidly with doping, from 10^{15} cm^{-3} in undoped a-Si:H up to more than 10^{18} cm^{-3} at the highest doping levels, as is shown in Fig. 5. These defects are charged, and therefore act as compensating centers that reduce the carrier density. This is one of the main reasons why the conductivity is lower than in crystalline silicon.

The second difference between doping in crystalline and amorphous silicon is that in a-Si:H the doping efficiency is very low. In the crystal, each impurity acts as a dopant, at least up to the solubility limit. In a-Si:H, most of the impurities are bonded to silicon (or hydrogen) in 3-fold coordinated sites which are inactive as dopants, and only a small fraction are 4-fold coordinated dopants. Furthermore this fraction decreases as the doping level is raised. In fact the density of active dopants is only slightly larger than the defect density, so that the compensation by the deep states is nearly complete. Fig. 5 shows that there is only a relatively small density of electrons which occupy band tail states and which provide the carriers for transport.

These unusual doping properties occur because the random bonding network in the amorphous film allows different bonding configurations from those in the crystal. The 3-fold coordinated state is the most energetically favorable for phosphorus or boron, and the 4-fold doping state has a low formation energy only when coupled with a charged defect. This situation can be described by the reaction for phosphorus doping,

$$P_3^{o} \leftrightarrows P_4^{+} + D^{-}$$

where the subscripts denote the coordination and the superscripts the charge, and D signifies the dangling bond defect.

A third distinguishing feature of doping in a-Si:H is that the doping efficiency depends on the thermal history of the material. Some examples

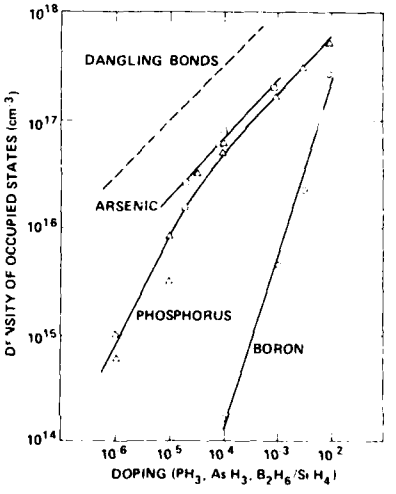


Fig. 5. Plot of the density of dangling bonds, and of shallow occupied band tail states for different doping levels. In undoped material the defect desity is about 10^{15} cm^{-3}

are shown in Fig. 4, in which it is seen that rapid quenching of the material from an elevated temperature increases the conductivity. The reason for this behavior is that the dopant and defect states are in thermal equilibrium at a sufficiently high temperature, as is discussed further in the next section.[13] Thermodynamics applied to the above doping reaction predicts that the defect density should increase as the square root of the doping concentration and that the doping efficiency should decrease in the same way. Both results are confirmed by experiment, and the square root dependence is shown in Fig. 5 The doping process in a-Si:H has proved to be fascinating aspect of the physics of amorphous semiconductors.

METASTABLE EFFECTS

One of the main material problems influencing the applications of a-Si:H in solar cells, displays and printers is the phenomenon of reversible metastable changes that are observed in the electronic properties. Such changes are induced by illumination, bias or an electric current, and are reversed by annealing to 150-200°C. The most widely studied metastability is the creation of defects by prolonged illumination, known as the Staebler-Wronski effect.[14] The defects created are dangling bonds and result from the recombination of electrons and holes within the material. As a result the solar cell efficiency decreases slowly with time, typically by 10-20% of its initial value. Much research has been devoted to understanding and eliminating these defects. A similar defect creation process often accompanies a shift of the Fermi (or quasi-Fermi) energy towards the band edge. For example, space charge limited current flow in p-i-p structures induces defects, and in thin film transistors, defects induced by the electron accumulation at the interface cause a shift of the threshold voltage.[15] The opposite process also occurs, in which a movement of E_F away from the band edges causes an improvement in the material properties. For example, annealing under reverse bias will improve the solar cell efficiency. Finally, metastability can be induced simply by thermal cycling, as discussed above and illustrated in Fig. 4.

The metastable effects are in fact closely related to the thermal equilibrium of the electronic states. The illumination or current flow etc., represents a departure from equilibrium, and as a result, defects are created. The equilibrium of a-Si:H is determined by the density of states distribution, and the characteristic time for the material to attain equilibrium. Experiments find that donor and dangling bond defect states come into equilibrium, and that the relaxation time is activated with an energy of about 1 eV. As a result of the high energy, equilibration takes over a year at room temperature, but only a few minutes at 200°C. The equilibrium state is understood in terms of the formation energies of the donors and dangling bonds and the above doping reaction. The formation energies depend directly on the position of the Fermi energy, E_F, since the defect and donor states are charged, and so a movement in E_F induces a change in the defect density. Our understanding of the metastable phenomena is therefore that the a-Si:H structure can be driven from its normal equilibrium by a shift of the Fermi energy, which can be by illumination, charge accumulation or temperature changes. The non-equilibrium state persists for a long time at temperatures below about 100°C, but is rapidly reversed at about 200°C. This process accounts for all the metastable effets.

The thermal equilibration of the electronic properties, and the metastable effects, are accompanied by structural changes. Experiments have concluded that the structural rearrangements are enabled by the motion of bonded hydrogen within the film. Evidence for the role of hydrogen is provided by the measurements of diffusion which is found to have an activation energy of 1.0-1.5 eV, similar to that of the electronic relaxation.[16] Furthermore, the observation of a dispersive time dependence of the hydrogen diffusion accounts quantitatively for the stretched exponential form of the relaxation of the electronic properties. Thus although hydrogen is generally beneficial for a-Si:H, it is also the underlying cause of the metastable effects. One of the challenges for research is to minimize the defect creation process.

DEVICE TECHNOLOGY

The particular electronic and structural properties of a-Si:H place certain design constraints on the devices that can be fabricated. For example, the high defect density in doped material results in a very short minority carrier lifetime. Thus solar cells are designed as p-i-n structures in which the p- and n-layers are made as thin as possible consistent with the formation of the junction, and the charge collection takes place entirely within the intrinsic layer. The doped layers are typically only 10 nanometer in thickness and the i-layer about 0.5 micron. In thicker films the loss of charge collection more that offsets any gain in absorbed light. The best way to increase efficiency is to grow multiple films of different band gap, and indeed the best devices are stacks of two or three cells. The thin film growth process makes it easy to fabricate such structures and materials of different band gap can be made, for example Si-Ge alloys by adding GeH_4 to the deposition gas. One active area of current research is the improvement in the material properties of these and other alloys, because at present their defect density is substantially higher than that of a-Si:H.

Similar considerations apply to transistor and

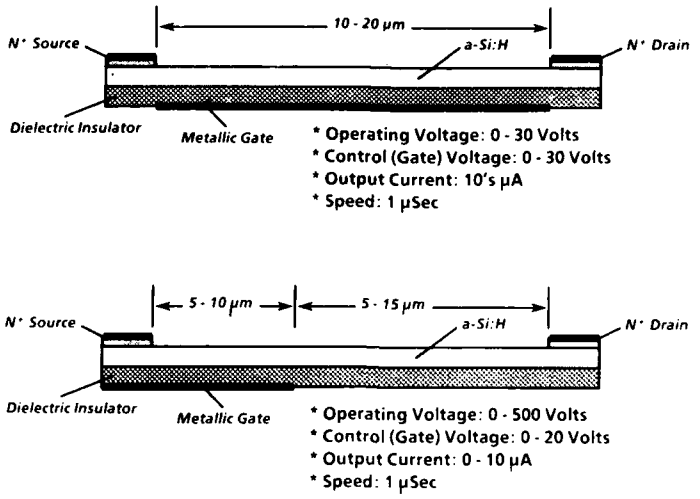


Fig. 6. Schematic diagram of the structure of planar low and high voltage a-Si:H thin film transistors, and their operating characteristics.

sensor structures. The active layer of a TFT is undoped a-Si:H, with n-type layers at the electrodes to give good ohmic contacts. The typical TFT structure and its operating characteristics are shown in Fig. 6. The TFT's are predominantly n-channel accumulation mode devices, because doped films cannot be depleted so easily, and p-channel devices have a low current because of the lower carrier mobility. The processing of a-Si:H circuits is in many respects identical to that for crystalline silicon, so that the photolithography and etching methods are well established. The need for large area devices has led to the development of processing capability of more that 12 inch square substrates with a feature size of 10-20 microns, and this will certainly improve in the future. The undoped a-Si:H layers, doped layers and gate dielectrics are all deposited by the same plasma process so that there is no need for dopant implants. Due to the diffusivity of hydrogen, a-Si:H films cannot be taken above about 350°C, so that thermal oxide dielectrics cannot be used. However the deposited nitride has proved to perform well with sufficiently few interface states or nitride traps. A curious result is that the interface state density is lower if the a-Si:H film is deposited on the nitride rather than vice versa, so that most TFT structures are inverted.

Although the current in planar TFT's is quite small because of the low carrier mobility, the thin film growth process makes possible vertical structures with a very short channel length and much higher currents, and such structures are being actively researched.[17] Another useful TFT device is shown in Fig. 6 and can switch up to 500 Volts with a low gate voltage.[18] The deposition process also allows for the formation of three dimensional electronic circuits which may be very useful in image processing applications.

Recent investigations have shown that a-Si:H films are also promising as detectors of high energy ionizing radiation.[19] The structures used are simple p-i-n devices in reverse bias. However the thickness needs to be much greater than with solar cells or TFT's to absorb the radiation. Devices of thickness at least 25-30 microns can be fully depleted, giving high collection efficiency. Because of the ability to make large arrays of devices and the fact that a-Si:H is more radiation hard than crystalline silicon, these films may have useful applications in radiation imaging.

SUMMARY

The research and device technology of a-Si:H is making very rapid progress. There is good control over material properties and much progress in the physical understanding of the material. An increasing variety of devices are being made, taking advantage of the large area capability and ease of processing. Solar cells, sensor and scanner arrays, and flat panel displays are already on the market, and many other products are under development.

REFERENCES

1. R. C. Chittick, J. H. Alexander and H. F. Sterling, J. Electrochem. Soc., 116, 77 (1969).

2. W. E. Spear and P. G. LeComber, Solid State Commun., 17, 1193 (1975).
3. D. E. Carlson and C. R. Wronski, Appl. Phys. Lett. 28, 671 (1976).
4. P.G. LeComber, W.E. Spear and A. Ghaith, Electronics Letters, 15, 179 (1979).
5. H. Miki, S. Kawamoto, T. Horikawa, H. Maejima, H. Sakamoto, M. Hayama and Y. Onishi, Mat. Res. Soc. Symp. Proc., 95, 431 (1987)
6. M.J. Thompson and H.C, Tuan, IEDM Tech. Digest, Los Angeles, 192 (1986).
7. R. A. Street, J. C. Knights and D. K. Biegelsen, Phys. Rev., B18, 1880 (1978).
8. R. A. Street, Proc. SPIE, 763, 10 (1987).
9. N. F. Mott and E. A. Davis, Electronic Processes in Non-crystalline Materials, Oxford University Press, 1979.
10. J. Kakalios and R. A. Street, Phys. Rev., B34, 6014.
11. M. H. Brodsky and R. S. Title, Phys. Rev. Lett., 23, 581 (1969).
12. R. A. Street, Phys. Rev. Lett., 49, 1187 (1982).
13. R. A. Street, M. Hack and W. B. Jackson, Phys, Rev. B37, 4209 (1988).
14. Staebler and C. Wronski, Appl. Phys. Lett.,31, 292 (1977).
15. C. van Berkel, and M. J. Powell, Appl. Phys. Lett. 51, 1094 (1987).
16. R. A. Street, C. C. Tsai, J. Kakalios, and W. B. Jackson, Philos. Mag. B56, 305 (1987).
17. M. Hack, J. G. Shaw and M. Shur, Proc. MRS Symposium on Amorphous Silicon, Reno, (1988) to be published.
18. H. C. Tuan, Mat. Res. Soc. Symp. Proc., 70, 651 (1986)
19. V. Perez-Mendez, S. N. Kaplan, W. Ward, S. Qureshi and R. A. Street, Nucl. Inst. and Meth. A260, 195 (1987).

THE IMPACT OF RAPID THERMAL PROCESSING

Thomas E. Seidel
Center for Robotic Systems in Microelectronics (CRSM)
University of California at Santa Barbara
Santa Barbara, California 93106, USA

Abstract

Rapid Thermal Processing (RTP) is an important furnace process technology. It has broad implications for the research and manufacture of advanced microelectronic materials and device structures. Advantages can be obtained through: annealing, or reactive film formation such as oxidation, nitridization, and silicidation, or rapid - chemical vapor deposition, or multistep and integrated processes. An overview of this technology is presented.

ADVANCED MICROELECTRONIC APPLICATIONS are being pursued using a flexible furnace technology, Rapid Thermal Processing (RTP). This paper discusses: the origins of the technology, demonstrations of the advantage of RTP, a list of applications, the new or emerging field of "Multiprocess and Process Tool Integration" using chemical vapor deposition CVD-RTP, and equipment issues. Typical heating and dwell times are several seconds for single wafers, while temperatures range from ~250-1400C. The use of higher temperatures combined with short time cycles allows for the selective advancement of thermal processes with different activitation energies.

ORIGINS OF THE TECHNOLOGY

Tungsten - Halogen lamps were used to anneal GaAs encapsulated in quartz by Kinsel in 1962.(1) The lamp approach was used to avoid contamination from the hot quartz walls of a standard furnace. In the mid-seventies, lasers were used as sources for the fast heating of implanted semiconductors, for the purpose of annealing crystal damage and to limit dopant diffusion.

RTP potentially covers a range of time durations from nanoseconds to about 100 seconds. The shortest time regimes are implemented with pulsed lasers and electron beams. For the shortest times, steep thermal gradients exist, with spatial and time temperature dependencies that follow the laws of thermal diffusivity. The shortest pulse durations provide adiabatic, abrupt thermal pulses to the surface region, while heating for intermediate time frames, (~10^{-5} to 0.1sec) gives a linear thermal gradient across the thickness of the wafer. Practical limitations preclude widespread use in silicon VLSI technology using very short time anneals. Thermal gradients stress the silicon and introduce either residual point defects or extended dislocations. A variety of optical interference effects occur near contact window edges and annealing is limited to the area of the laser exposure, so the process has low thruput. Technical aspects of the shorter time RTP are addressed by Hill.(2) Annealing with times above ~1 second give nearly isothermal wafer heating and it is this case has more widespread practical applications.

THERMAL COUPLING, GRADIENTS

A variety of heating sources can be used to obtain an isothermal heating of single wafers. Arc lamps, tungsten-halogen filaments, or graphite resistive elements can be used. Lamp heating rates can differ from ~0.3-10 sec, but the heating rate of the wafer is determined by its optical coupling, heat capacity and wafer thickness. The time to heat the wafer (3) is typically the order of a few seconds.

$$dT/dt \sim (1-R)\int_{\lambda} d\lambda\, I_{\lambda}(1-\exp(a_{\lambda} d)/C \qquad (1),$$

where R is the reflectivity, C the heat capacity, a the absorption coefficient, I the incident energy intensity. (a, I and R are dependent on the irradiating wavelength), and d is the thickness of the wafer. The thermal conductivity / diffusivity of silicon is large enough so that it is difficult to obtain a very large thermal gradient *across* the thickness of a wafer, even if the wafer is heated from only one side.

Celler (4) has demonstrated vertical gradients of a ~7 degrees for the case of silicon heated by radiation from one side near the melting point of silicon.

$$dT/dz = eST^4/K \qquad (2),$$

where e is the emissivity, S=5.6x10^{-8} $Wm^{-2}K^{-4}$ is the Stephan-Boltzmann constant, K the thermal conductivity, z the depth coordinate and T the temperature. In this case silicon melts at the hotter surface and a "mixed state" of molten islands are formed in a background sea of crystalline material. The reflectivity of molten silicon is close to unity, as it is for

metallic silicon phase. As the power is increased, the molten islands grow in size but the average temperature of the wafer system remains essentially constant as the energy goes into melting the entire surface. It is possible to hold the hotter surface at 1412C and the colder surface-device side at 1405C for an indefinite period without any electronic feedback.

Under some RTP conditions a significant *lateral* or radial gradient exists. The temperature can be considerably different in the central region of the wafer than the edge, depending on radiative conditions. The temperature gradient can lead to stresses above the yield stress and result in "slip" dislocations. However, for the most cases, we consider the wafers to be isothermal.

RTA ADVANTAGE, DEMONSTRATIONS

One might think that temperature can be traded for time for thermally activated processes and after some period in the kinetics, an equivalent annealed - physical state would be obtained. This is true if we are considering a single thermally activated process, and not so if we have two or more competing processes with different activation energies. At higher temperatures, the higher energy process proceeds relatively more quickly. For implanted dopants in silicon, the activation energy for damage removal is 5eV and the dopant diffusion energy (generation plus migration) is ~3.5eV. Thus the damage can be removed at high temperatures while limiting the diffusion. This has been illustrated by Sadana, et al. (5) for boron implants and Seidel, et al. (3) for arsenic implants. The concept is not limited to implant activation applications. In metallurgical applications, phase formation with limited diffusion is often desired.

In Fig.1 boron profiles are compared for RTA and standard furnace anneals. The anneal conditions are adjusted so the junction depths are the same but the damage is different. The RTA case has a broadened peak, while the furnace case has a pronounced peak and residual dislocation damage.

In Fig.2, dopant profiles are plotted for the cases where dislocations are first removed as time is increased at each temperature. For example, dislocations still exist at 8sec but not at 10sec at 1100C. The dislocations at issue are the localized dislocation defect arrays associated with the damage for arsenic ions stopping in the crystalline material just beyond the amorphous/crystalline (a/c) interface depth position. See arrow. Clearly the defect free profiles are more limited in their diffusion for the higher temperature cases.

Another example demonstration case is the limited diffusion of silicon in aluminum when the contact reaction sintering between silicon and aluminum is carried out at ~450C for a few seconds, see Pai, et al., (6).

Also, thin thermal-nitridizations are done for short times at very high temperatures, ~1150C, where the thermal diffusivity of nitrogen through the thin underlying oxide and the accumulation of nitrogen at the silicon dioxide/silicon interface is limited. The activation energy for the nitride formation is greater than the energy for diffusion. See Chang, et al., (7).

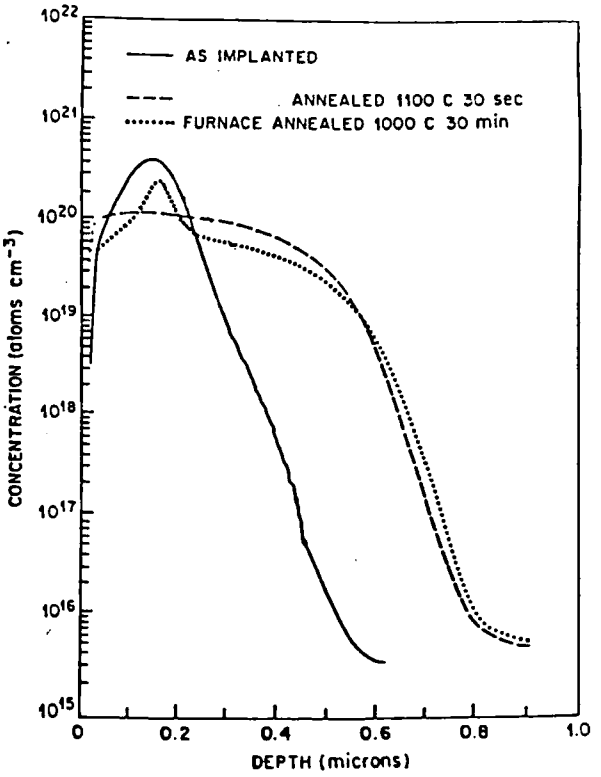


Fig.1. Concentration profiles for boron implanted into crystalline silicon. The standard furnace annealed sample has a severe residual dislocation damage. After Sadana, et al, (5).

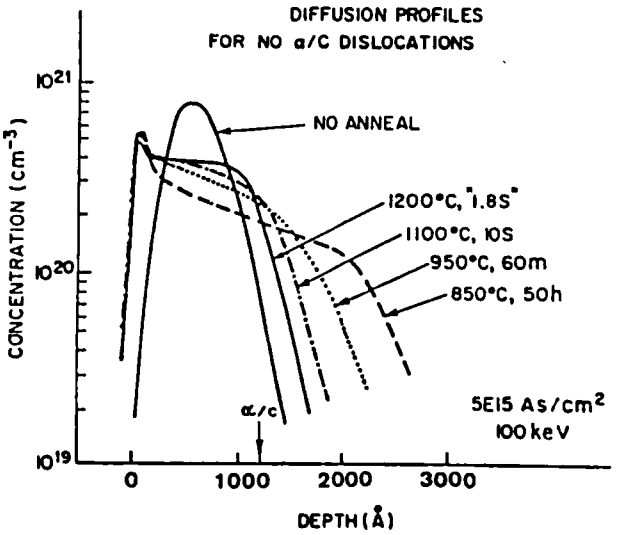


Fig.2. Concentration profiles of arsenic implanted silicon. The time was adjusted at each temperature until the residual dislocations in the space charge edge just beyond the a/c marker were removed. After Seidel, et al, (3)

APPLICATIONS

Applications also address issues other than the utilization of the "RTA advantage" where multiple, competiting thermally activated processes are the issue. Multistep process integration may drive the applications.

REACTIVE FILM FORMATION - Rapid Thermal Oxidation (RTO) and Nitridation (RTN): Thin gate oxides with quite high dielectric breakdown strength and good uniformity have been formed by RTO. The oxide quality is at least as good as that obtained using conventional furnaces. There is discussion over the possibility of fundamentally different oxide interfaces and the claim that SiO_2 interfaces are atomically smoother because of the higher temperature of formation and "flowed" oxide conditions. Film uniformity of ~2%, over most of the wafer has been demonstrated. Perhaps the most important aspect of RTO is the process compatibility with idea that multistep processes and process tool integration can be utilized for the next steps in the process flow.

The next step may be rapid thermal nitridation (RTN), See Nulman (8) Fig. 3, and/or deposition of CVD polysilicon, and/or the deposition of CVD metal for silicides. *It is this flexibility and integration of the multistep processes that has potential commercial importance for the RTP approach.* The chamber in which the processes are being carried out have high ambient integrity, and process ambients can be changed in times comparable to the short process time. The strategy would be to run chemically compatible multistep processes in the same chamber and add separate chambers for the next increment of process integration.

SILICIDE FORMATION - The ability to react silicides in short times at elevated temperatures and controlled ambients favors the use of RTA for certain silicides.(9) Titanium is extremely reactive with oxygen and requires low oxygen for the controlled reaction with silicon to proceed. (10) The reaction proceeds by the diffusion of silicon into the metal. This causes the silicide to form on the sidewalls of oxide-cut topology in addition to forming TiSi in junction contact regions. (In standard furnaces, oxygen can back stream into the wafer area.) TiSi reactions are an important example of a process which is better done using a high ambient integrity RTP/RTA approach. (11).

The diffusion of silicon during the silicide formation results in the liberation of vacancies. The excess vacancies participate in the enhanced dissolution of dislocations. See Rozgonyi, (12). The dislocations in silicon are generally composed of excess silicon intersitials and thus are dissolved more readily by a flux of excess vacancies.

SHALLOW JUNCTIONS - The idea that shallow junctions can be produced by low energy implantation into crystalline silicon is thwarted by the variability of ion channeling effects. Implanted boron dopant ions undergo wide angle scattering into the open channels of the crystal target, creating penetrating tails on the profile of the implanted distribution. Also, the stopping boron ions create damage clusters which lead to low temperature enhanced thermal diffusion.

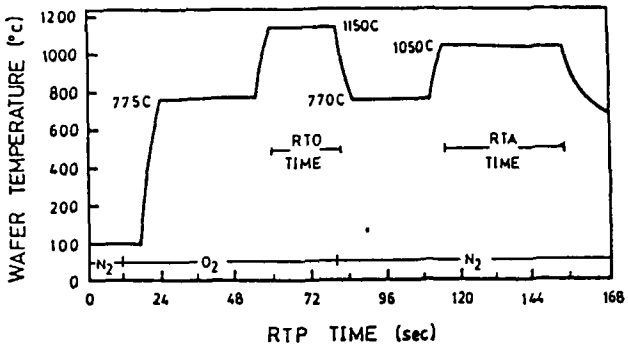


Fig.3. Temperature of wafer plotted against anneal cycle time, for a typical rapid oxidation / anneal (N_2) or thermal nitridation (NH_3) sequence. After Nulman, (8).

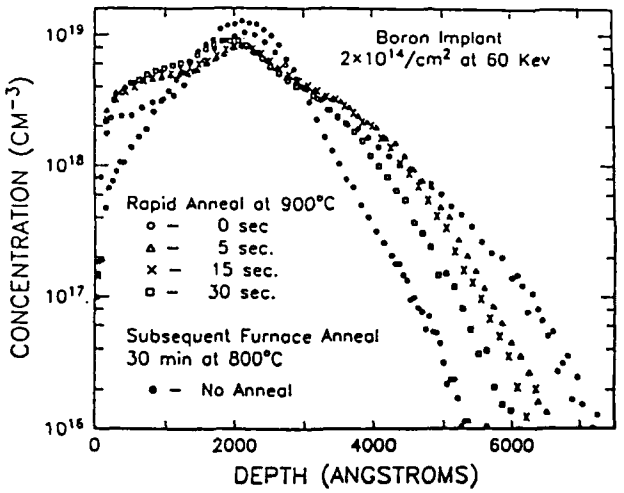


Fig.4. Boron SIMS profiles with a 900C RTP preanneal, before a 800C-30min furnace anneal. After Michel, (13)

The channeled tail of boron exhibits a several order of magnitude enhanced thermal diffusion. See Michel, (13, 14), and Fig. 4. The enhanced diffusion at 800C has been reduced by an RTA preanneal at higher temperatures. The detailed mechanisms for the enhanced diffusion are now unclear, but it is possible that the boron and local damage result in an interstitial-like enhanced diffusion process, while it may also be possible that point defects come from sources that are non-local to the channel tail region. Interstitial silicon point defects are known to replace substitutional boron dopant and enhance the boron diffusivity as long as the boron is interstitial.

Another approach has been the use of preamorphized silicon by the pre-implantation of an "inert" damaging species, such as silicon implanted into silicon, (3) When sufficient damage occurs the crystal disorganizes into an amorphous state and no channeling is possible for subsequent ions implanted into this material. However, there is dislocation damage just beyond the amorphous / crystalline (a/c) interface which should be annealed out for low leakage junctions to be obtained.

The implanted distribution can be contained entirely within the thickness of the preamorphized layer, and then diffused into crystalline material. The dopant impurities may be diffused beyond the a/c disorder and this gives quality junctions with a depth of about 0.1um. The shallowness of the process is limited because the width of the residual dislocation a/c array is 500A wide, and the minimum available practical energy of the implant is ~5keV (range of 500A).

One may also consider implantation of dopants into film-like layers above the crystalline silicon followed by out-diffusion into the crystalline silicon. (15). In this manner there may be shallow, controlled junction profiles, without channeling and ion damage effects in the crystalline silicon.

The junction technology "process - package" is dependent on many issues. The specific application and process compatibility with the entire IC process needs to be considered. Issues such as contact resistance, high-electric field effects, and glass flow for sodium passivation are as important as the the junction depth and the control of MOS channel lengths.

The annealing of implanted III-V compounds for the purpose of obtaining shallow junctions has been done out of necessity by RTP for some time. Among other advantages, the short thermal cycles limit the evaporation of group V elements from the surface, see Streetman (16).

CHEMICAL VAPOR DEPOSITION - CVD/RTP under low pressure conditions – has been used to grow thin films, resulting in the controlled deposition of extemely thin heavily doped silicon films by Gibbons, et al.,(17). See Fig. 5. Basically, the temperature is switched high for deposition and switched low to turn off the thermally activated deposition cycle.

System features include many advantages: a closed system with ambient integrity allowing for purging, insitu cleans, and fast gas switching and control, a cold wall which promotes furnace cleanliness, and independently controlled temperature - time and process gas flow - time recipes. The flexibility of the approach far exceeds that of conventional batch loaded furnaces, where access, ambient control, and speed of temperature changes are limited.

The abiltity to perform RTP/CVD now includes the following films:

a) Silicon Epitaxy, thin doped layers.
b) Doped and Undoped SiO2 Glass.
c) Polysilicon, doped and undoped.
d) Tungsten Doposition.

Oher metal silicides can be pursued: Ti, Co and Cu. Various III-V compounds can be deposited (17,18).

In addition, as new chemicals are developed, new superconductor thin film may be deposited using RTP/CVD, where it is very important to limit the diffusion of imputities from the substrate.

OTHER APPLICATIONS - The use of implantation in IC manufacture often requires the assessment of the dose and uniformity of the implant as soon as possible after the implant has been made. This is currently one of the most common uses of RTP equipment. Implant dose and dopant diagnostics can be made immediately, resulting in high confidence in wafer product and implant equipment. See Keenan (19).

Gettering of impurities and defects away from the active regions of devices have been important for obtaining large area wafer perfection, which allows the advance of the level of complexity in IC technology. The ability to release impurities from dislocations via the dissolution of dislocations seems to be a promising use of RTA, see Sparks (20).

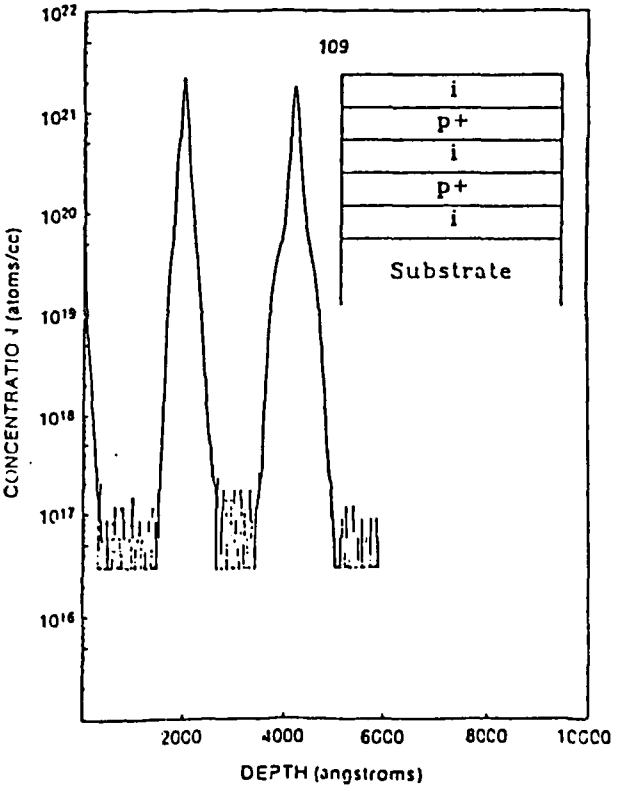


Fig.5. Concentration profiles for boron, formed using thermally switched RTP/CVD or "Limited Reaction Processing" After Gibbons, et al. (17)

PROCESS TOOL INTEGRATION

We can anticipate further process integration for microelectronic manufacture as the capability for multiprocess deposition tools continues to develop. Facility architectures that enhance yield, cycle time and thruput are emerging. One design concept is to place small footprint process tools locally together that are used serially in the process flow sequence within a single clean envelope area, and connect them by automated wafer transfer mechanisms. An example facility architecture has beed proposed, (21) It is a Self - Contained - Automatic - Robotic - Facility or "SCARF". Considerable improvements in cycle-time may be obtained, especially if sensor technology is applied to the control and monitoring the processes. An example layout of a multiprocess deposition and etching sectors is shown in Fig. 6.

EQUIPMENT

The basic heating mechanism for wafers includes the overlap of wafer absorption and incident "black body" spectrum. The wafer system, depending on the films near the surface and the doping (infrared free carrier absorption) will have a different optical response to the incoming radiation. The success of future RTP furnace operations is dependent on the ability to heat the wafers with very good lateral temperature *uniformity*, and the ability to measure the *temperature*. Knowledge of the temperature has been a concern since the earliest days of the technology.

Thermocouples are good for experimental purposes but are destructive for integrated circuit wafers. Optical pyrometry is an excellent approach for the case of a wafer without dielectric layers, but most of the important applications use wafers with multi-layered dielectric and metal structures. More fundamental methods for measuring the temperatue are needed, measurement of the lattice constant, expansion coefficient or the line-width of temperature sensitive optical emissions would be examples of more direct temperature metrology.

Other equipment issues are: monitoring the gas flow and gas chemistry (fluoresence), pressure, cold wall furnace temperature and cleanliness control, power - time controlling algothrims. Both temperature and gas delivery uniformity impact the film deposition uniformity. See Gelpey (22) The design of the wafer edge environment is crucial for the control of a lateral or radial temperature uniformity although some control is also afforded by modifying the distribution of incident radiation. More energy can be delivered to the edge region of the wafer than the center. Process control using flexible recipes are available.

Finally, as the degree of process integration and complexity increases, the requirements on process tool *reliability* and cleanliness become crucial.

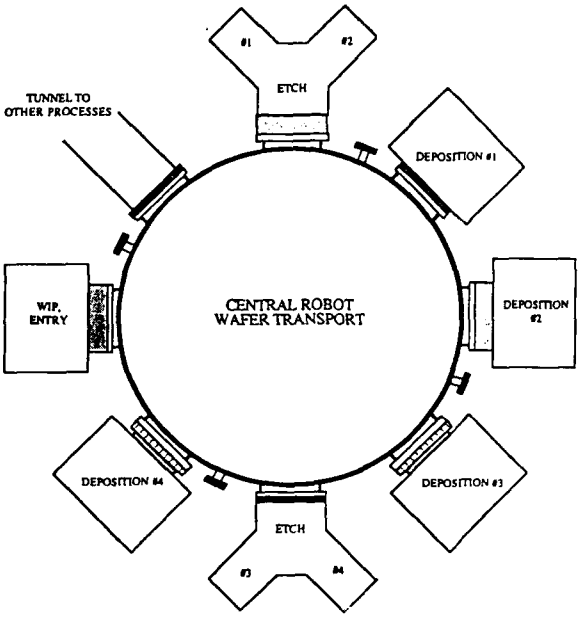


Fig 6. Schematic of an advanced integrated process system, the modules and central chamber are self-contained. After Skidmore, (21)

ACKNOWLEDGEMENTS

I would like to thank E. Hu, D. Maher and S. Shatas for interesting and valuable discussions.

REFERENCES

1. Kinsel, T. S. and T. E. Seidel, J. Appl. Phys. 33, 767 (1962)
2. C. Hill, in "Laser and Electron Beam Solid Interaction and Material Processing", p.361, J. F. Gibbons, L. D. Hess and T. W. Sigmon, Eds., North Holland, New York,, (1981)
3. Seidel, T. E., D. J. Lischner, C. S. Pai, R. B. Knoll, D. M. Maher, and D. C. Jacobsen, Nucl. Inst. & Meth. in Phys. Res., B7(8), 251 (1985)
4. Celler, G. K. Solid State Technology, 30 (3), 93 March (1987)
5. Sadana, D. K., S.C. Shatas and A. Gat, , Inst. Phys. Conf. Ser., 67, 143 (1983)
6. Pai, C. S., E. Caberos, S. S. Lau, T. E. Seidel, and I. Suni, Appl Phys Lett. 46 (7), 652 (1985).
7. Chang C., A. Kamgar, and D. Kahng, , IEEE Electr. Dev. Lett. EDL-6 (9), 476 (1985)
8. J. Nulman, "Ext. Abstr. The Electrochem. Soc. 88-1, No. 63 (1988)
9. Shih D. K., S.K. Lee, Y.H. Ku, D.L. Kwong, and S. Lee, Mat. Res Soc., Symp. 92 ,121 (1987)

10. S. P. Murarka, "Silicides for VLSI Applications", Academic Press, New York, 1983.
11. J. Pramanik, "Semiconductor International", p. 94 Cahners Publ. Co, Newton MA, May (1985)
12. Rozgonyi G. A., Ext. Abs. The Electrochem. Soc. 88-1, No 83 (1988)
13. A. E. Michel, in "Rapid Themal Processing", T. O. Sedgwick, T.E. Seidel, and B-Y Tsaur, Eds, , Matl. Res. Soc., Pittsburgh, PA(1986)
14. Michel A. E., Ext. Abstr., The Electrochem. Soc., vol. 88-1, No. 72 (1988)
15. F-C Shone, K. C. Saraswat and J. D. Plummer, Stanford Electronics; Lab. Report, ICL17-79, (1987)
16. Streetman B. G., and A. Dodabalapur, Ext. Abstr., The Electrochem Soc., 88-1, Abs. No. 78., (1988)
17. Gibbons J. F., S. Reynolds, C. Gronet, D. Vook, C. King, W. Opyd, S. Wilson, C. Nauka, G. Reid, and R. Hull, Mat. Res. Soc. 92, 281, (1987)
18. Vook, D. W., S. Reynolds, and J.F. Gibbons, Appl. Phys. Lett. 50 (19), 1386 (1987)
19. Keenan, W. A., Ext. Abstr., The Electrochem. Soc., 88-1, No 85, (1988)
17. Sparks D. R., R. G. Chapman, N. S. Alvi, Appl. Phys. Lett. 49, 525 (1986)
21. Seidel, T. E., Ext. Abstr., The Electrochem. Soc., 88-1, Abstr. No 71., (1988), and K. Skidmore, "Semiconductor International", p79 Cahners Publ. Co. Newton MA, May 1988
22. Gelpey J. C., Ext. Abstr., The Electrochem. Soc., 88-1, Abstr. No. 84., (1988)

PRINCIPLES OF STM AND APPLICATIONS IN THE SEMICONDUCTOR INDUSTRY

M. D. Pashley
Philips Laboratories
Briarcliff Manor, New York, USA

D. A. Smith
IBM Research Division
Yorktown Heights, New York, USA

ABSTRACT

Scanning Tunneling Microscopy (STM) has quickly become established as an atomic scale characterization technique which is complementary to high resolution transmission electron microscopy and field ion microscopy. The STM has successfully been applied to both atomic and electronic surface structure and new results show great promise in the study of molecular beam epitaxial growth. Additional applications of STM extend to the nano-scale manipulation of surfaces by deformation and lithography.

THE SCANNING TUNNELING MICROSCOPE invented by Binnig and Rohrer [1] has rapidly achieved an impressively broad application, recognition and acceptance in the scientific community. Even though new applications of the technique continue to appear it is possible to offer a perspective on the utility of the scanning tunneling microscope (STM) and to compare it to alternate techniques. The main focus of the present overview is on the role of STM in the characterization of the materials and processes which are of particular importance in the electronics industry.

OBTAINING THE IMAGE

The STM obtains information from a sample through the passage of a tunneling current between regions of atomic dimensions on a probe and the sample (Fig. 1). The production of suitable probes is a mixture of art and science. It is remarkable that the early mechanical means of probe fabrication were successful. Presently probes are made by chemical or electrochemical polishing of axially symmetrical blanks using techniques developed in field emission and ion microscopy where the sample must be in the form of a sharp point. Probes can be made from virtually any metal but are usually tungsten. The tunneling current i depends on the potential V applied between the probe and the sample, the effective work function ϕ and the separation d between the probe and sample as described below

$$i \alpha f(V) e^{-A\phi^{1/2}d}$$

where A is a constant. The exponential dependence of tunneling current on separation is exploited in the STM. The probe is mounted on piezoelectric elements which permit the motion of the probe in an x-y scanning fashion and also in the z-direction perpendicular

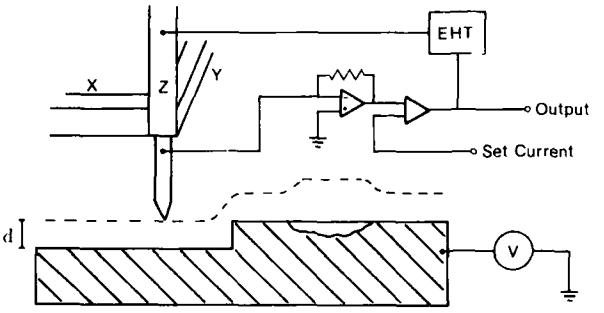


Figure 1: Schematic of the principle of operation of the STM with the tip attached to x, y and z piezos at a distance d from a surface. The dashed line shows the STM contour of constant current over a step and a region with different electronic characteristics - either a reduced work function or increased electron density of states.

to the sample. With other factors constant when the probe is scanned across the sample, the tunneling current depends on d. Thus if d is constant so is i. With a voltage of 1 volt and a typical tunnel current of 1nA the separation is in the 5 - 10 Å range. The key to the original operating mode of the STM is to maintain the tunneling current constant by compensating any change in d resulting from the specimen topography by a corresponding *motion of the probe* by changing the potential applied to the z-piezo (see Fig. 1). This potential is the signal which is amplified and processed to provide an image. This mode of operation is known as the constant current or "topographic" mode. Data are acquired relatively slowly in this mode (typically 100 - 500 data points per second), the rate of acquisition being limited by the time constant of the feedback loop which controls the z-piezo. For atomically flat samples the probe may be scanned at such a rate that the z-piezo feedback circuit cannot respond in which case the tunneling current does vary and this variation is a direct indicator of the surface structure. This is called the constant height mode. It has the advantage of speed but does not directly give quantitative height information. A third operating mode which gives local electronic information is to monitor the tunneling current at constant x, y and z whilst changing the potential between the probe and the sample.

The STM has proved able to operate in diverse environments which include UHV, conventional high vacuum, air and various liquids. Frequently an STM is integrated with other surface characterization tools and facilities for in situ processing. The present trend is to configure the STM specimen holder to permit in situ processing such as heating, deposition, ion bombardment and implantation together with characterization of the probe and the sample by means such as RHEED and LEED, AES, ESCA, FIM, SEM and REM.

DISPLAYING THE IMAGE

The output data from the STM is in the form of a height (or current in constant height mode) for each x-y point in the image. This can be displayed in various ways. The simplest method is to plot y+z as a function of x producing a linescan image. An example of this is shown in Fig. 2a for a 70Å x 60Å area of a Si(111) - (7x7) surface with 128 x 80 points in the image (data from ref. [2]). This type of display has the advantage of simplicity in that it only requires a pen recorder or oscilloscope. However, it is not the most informative method of display. A line scan image can be converted into a grayscale map on a CRT with white high and black low as seen in Fig. 2b, where the only processing is scaling to fit the grayscale range.

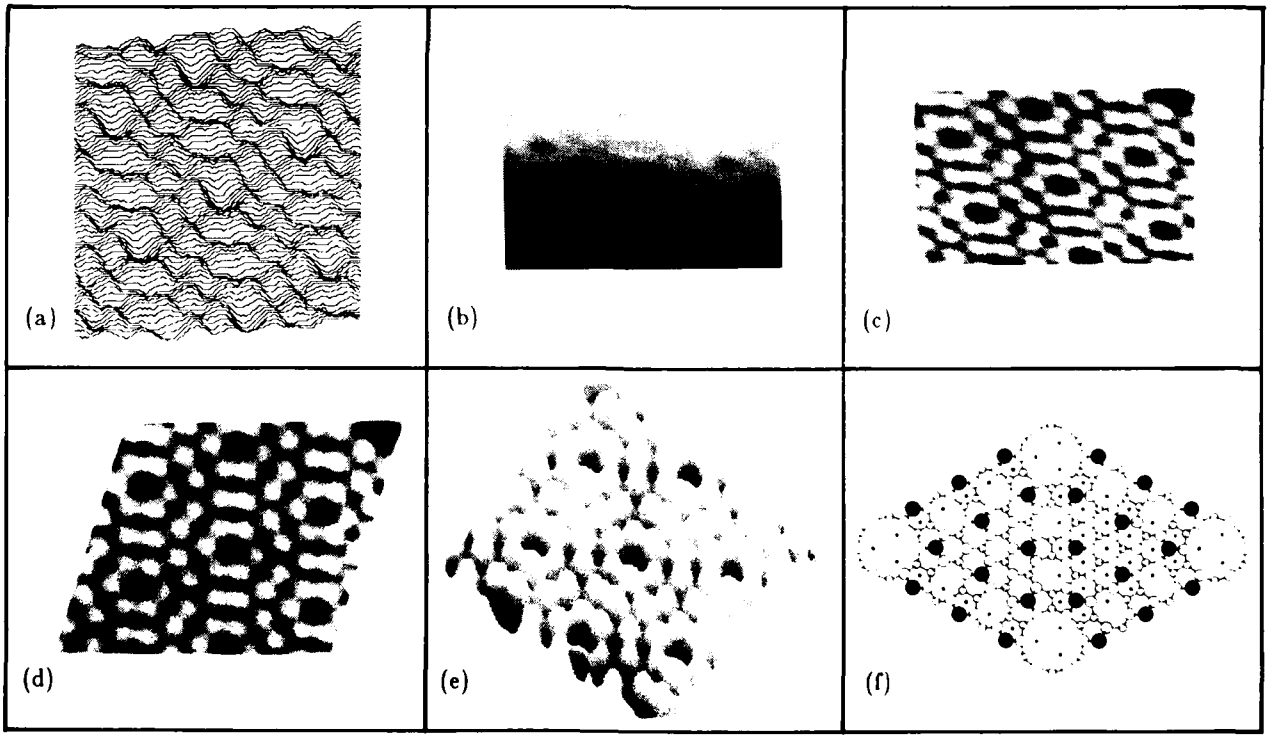


Figure 2: An image of a 70Å x 60Å area of the Si(111) - (7x7) reconstruction (data from Ref. [2]): (a) linescan, (b) grayscale image, (c) after plane subtraction, (d) after drift correction, (e) 3D perspective view, and (f) DAS model of the (7x7) reconstruction with the large filled cicles representing the adatoms imaged by the STM (after ref. [10]).

Frequently the image will have a slope to it arising from tilt on the specimen or drift of the z piezo, which dominates the display. This can be removed by subtracting a background plane from the data calculated with a least squares fit to give Fig. 2c. The least squares fit can only be done after data acquisition is complete. For real time plane subtraction, approximations using only the current line can be used. The structure is now clearly seen, but due to thermal drift the expected symmetry is not seen. Thermal drift is most easily determined by comparing the xy coordinates of a particular feature on two consecutive images [2]. In addition to thermal drift there may be distortions to the image resulting from cross-talk between the x and y piezo. This can be determined by imaging a known structure with high symmetry such as the Si(111) - (7x7). If after drift correction the correct symmetry is not obtained then the remaining distortion arises from such instrumental effects and can be determined and then used as a general correction factor. The result of xy cross talk and thermal drift correction are shown in Fig. 2d. This is generally the most useful way of displaying STM data. In this case the features seen are the surface adatom structure (see below for further discussion). Further processing is often not required although a low pass filter can usefully remove noise. Other more complex filtering techniques can be helpful with noisier images [3]. However, stepped surfaces are often not displayed successfully by this technique. Instead a combination of grayscale and linescan display can be used producing the 3D perspective plot shown in Fig. 2e. The linescan is rotated and tilted in order to give the required perspective view and the gray levels are determined from the height at each point as with the direct grayscale image. It is also possible to determine the gray levels from curvature in order to enhance surface features.

IMAGE INTERPRETATION

The image obtained from the STM contains both atomic and electronic information. The change in atomic height associated with a step is clearly seen. However the features seen within an atomically flat plane can be more complex. The theory of Tersoff and Hamann [4] shows that the STM profile is closely related to contours of constant charge density. Thus by varying the tunnel voltage, and hence the energy states of the surface available in the tunneling process, the spatial distribution of particular electron energy states can be probed. This is very well demonstrated by the work of Feenstra et al [5] on the GaAs(110) surface. The surface structure has one arsenic and one gallium atom per unit cell. However the STM image obtained at +1.9V shown in Fig. 3a shows a surface with only one feature per unit cell. The corresponding

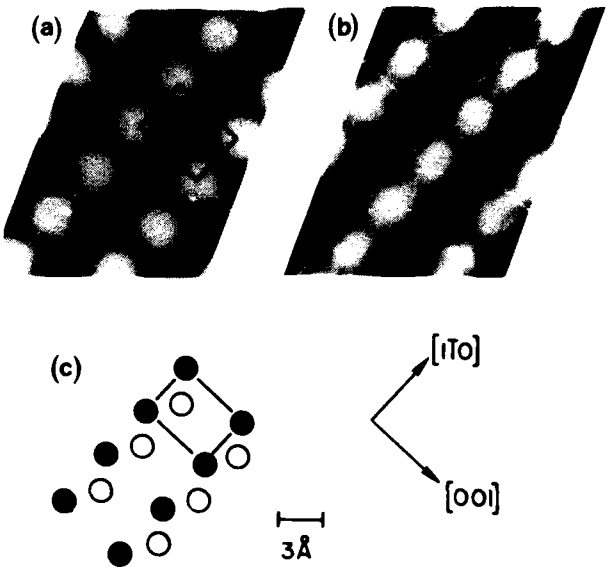


Figure 3: Constant current STM images from ref. [5] of the GaAs(110) surface showing the distribution of (a) empty states and (b) filled states. The images were taken at +1.9V and -1.9V respectively. The top view of the surface atoms is shown in (c) with the As atoms represented by open circles and the Ga atoms represented by closed circles. The rectangle indicates a unit cell whose position is the same in all three figures (courtesy APS).

image taken at -1.9V (Fig. 3b) is very similar, except that if the two are overlaid, the bright features of the two images do not coincide. This is shown by the rectangle which indicates a unit cell in the same position in both images. Images taken at positive polarity reflect the distribution of empty states on the surface (electrons tunneling from the tip to the surface) whereas images taken at negative polarity reflect the distribution of filled states on the surface. In the case of the polar semiconductor gallium arsenide, filled states are localized on the arsenic atoms while empty states are localized on the gallium atoms. Thus the images shown in Fig. 3a and 3b are images of the arsenic and gallium atoms respectively. We therefore see that by taking images at different voltages or polarities much more information on the surface structure can be obtained than is possible with one image, and that in fact a single image can be misleading.

The study of the electronic structure of the surface can be extended by taking I-V spectra at a fixed point on the surface [6]. This is achieved by holding the tip stationary at the point of interest on the surface,

momentarily disconnecting the feedback loop with the tip to sample separation held constant, and sweeping the tunnel voltage while recording the tunnel current. The current as a function of voltage gives information on the local density of states [4] with the ability to study both occupied and unoccupied states (somewhat analogous to a combination of photoemission and inverse photoemission). This type of measurement has been combined with scanning in Current Imaging Tunneling Spectroscopy (CITS) by Hamers et al [7]. Here the feedback loop is disconnected at each point in the image and an I-V curve taken. Then in addition to the "topographic" image, an image of the current at several specified tunnel voltages can be obtained. Provided that the images are made at voltages corresponding to strong features in the density of states, the spatial distribution of a particular surface state can be mapped out. The different techniques for voltage dependent STM imaging are discussed in more detail in ref. [8].

THE Si(111) SURFACE

The Si(111) - (7x7) reconstruction was the first semiconductor surface to be imaged by STM [9]. These data reduced the number of valid models for the structure of this surface, and gave strong support to an adatom structure as the outer layer. However, the STM only images the local charge density at the surface, and so does not directly give information on layers below the surface which may also be involved in the reconstruction. The structure of the (7x7) reconstruction was finally determined from Transmission Electron Diffraction [10]. The resulting Dimer Adatom Stacking Fault model (DAS) [14] is shown schematically in Fig. 2f. The 7x7 periodicity results from a stacking fault in one half of the unit cell, in the double layer below the surface adatom structure. It is only this outer adatom layer that is directly imaged by the STM. Studies of the STM image as a function of voltage, together with detailed comparisons of the STM linescans obtained from experiment and those calculated for various models proposed to be consistent with the original STM images of Binnig et al [9], have shown the DAS model to be the best fit [11]. The spatial location of surface states on the (7x7) surface has been imaged by CITS giving further evidence in support of the DAS model [7]. Thus although the determination of the full structure of the Si(111) - (7x7) surface was not made with the STM, the results obtained with it were important in narrowing down the possible models. With the further development of spectroscopic techniques and theoretical calculations the ability of the STM to distinguish between competing structural models is increasing.

MBE GROWTH

The understanding of the surface processes which are involved in film growth is central to the production of defect free materials suitable for electronic applications. It was realized at an early stage of the development of materials science that migration to and addition to steps were likely to be the fundamental processes in growth from the vapor or solutions. Only recently have high resolution electron microscopy and STM made possible direct observation of these processes. The STM permits the topography of both the substrate and the deposit to be examined. It is here that the high resolution real space imaging capability of the STM is uniquely able to image the defect and aperiodic structures on the surface which are so important in growth, as well as image surface reconstructions. We now discuss two recent STM studies that are

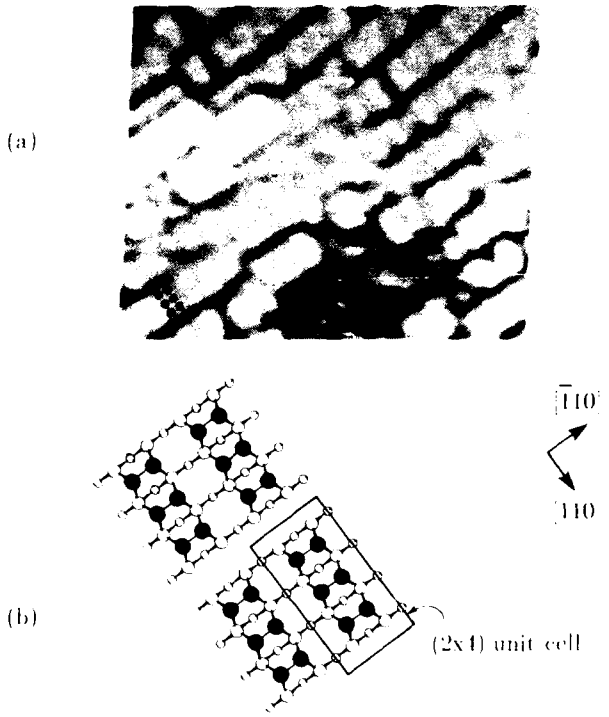


Figure 4: (a) An STM image from ref. [12] of the GaAs(001) (2x4) surface showing a 165Å x 130Å area including small islands one plane up (light) and one plane down (dark), each having the same structure as the main plane. The atom positions are marked with filled circles and the missing atom sites are marked with open circles. The image was taken with the sample at -2.3V and the grayscale covers a range of 7Å. (b) Diagram of the structure of the (2x4) unit cell with the surface As atoms represented by filled circles, the top Ga layer represented by large open circles and the 2nd As layer represented by small open circles.

related to important issues in Molecular Beam Epitaxy (MBE) growth.

The (001) surface of GaAs is important in device growth. MBE growth is usually carried out under arsenic rich conditions and so it is the arsenic rich surface which is most important in understanding the growth mechanisms. This surface was known to form a (2x4) reconstruction, but the structure of the unit cell was not understood. Recent STM results of Pashley et al. on an MBE grown GaAs(001) surface [12,13] have shown that the (2x4) reconstruction arises from the formation of arsenic dimers; every 4th dimer in the [110] direction is missing, consistent with a model proposed by Chadi [14]. A typical image of this surface is shown in Fig. 4a. The surface structure is shown schematically in Fig. 4b. In this case, the individual dimers were not generally resolved. The dark rows of missing dimers running along the $[\bar{1}10]$ direction are clearly seen together with the 2x periodicity arising from the arsenic dimers. In addition to determining the surface structure, the STM results have shown for the first time that the unit cell of a block of 3 arsenic dimers forms the basic building block for the structure of islands and steps on this surface. The islands are made up from complete unit cells and can be only one unit cell wide in the 4x direction (see Fig. 4), with step edges tending to run along the $[\bar{1}10]$ direction. Without this type of detailed structural information, atomistic models of growth mechanisms cannot be developed.

There is considerable interest in growing GaAs on Si by MBE in order to make large scale GaAs devices on readily available large and inexpensive Si wafers. The quality of the grown GaAs layer is highly dependent on the structure of the Si substrate. The presence of single atomic height steps on Si(001) is believed to result in the formation of antiphase boundaries in the GaAs (see Fig. 5a and b), whereas double steps do not. The STM provides an ideal way of studying the structure and distribution of the steps on Si(001) substrates. An STM image of an array of double height steps obtained by Wierenga et al [15] is shown in Fig. 5c. These were observed on a substrate cut 4° off the (001) plane about the $[\bar{1}10]$ axis. Growth of GaAs on this surface produces a low density of antiphase boundaries. In contrast, a Ge(001) surface (which is known to behave similarly to Si(001)) cut 2° off the (001) plane showed predominantly single height steps [16,17]. The direction of the tilt axis can also have a significant effect on the step structure on the surface. This work has already illustrated how preparation of the substrate can have a major effect on the structure of steps and hence on the quality of layers grown on the substrate.

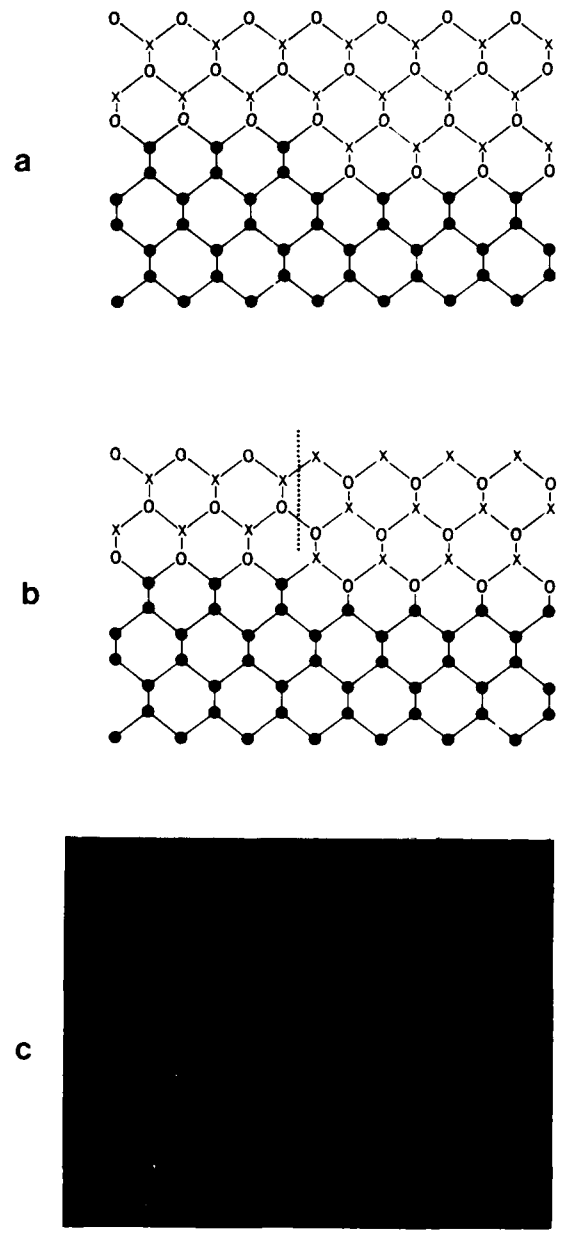


Figure 5: The growth of GaAs on Si(100) with (a) a double step showing that the correct ordering is obtained in the GaAs layer and (b) with a single step showing the formation of the antiphase boundary indicated by the dotted line. The Si atoms are represented by filled circles, the As atoms by open circles and the Ga atoms by crosses. (c) STM image from ref. [15] of biatomic steps on the vicinal Si(001) surface. The image shows a 150Å x 150Å area and was obtained with the sample at -2V (courtesy APS).

These two STM studies clearly demonstrate that in the field of growth, the STM can directly address issues of immediate importance. Apart from one other recent study of the initial stages of epitaxial growth of As on Si(111) which is really the first stage in the growth of GaAs on that surface [18], there has been very little other STM work on MBE growth related issues. However a marked increase in the STM study of systems important in MBE growth can be expected.

THE REACTION OF PALLADIUM WITH SILICON

The formation of silicides by reaction between a deposited metal and a silicon substrate has become a key part of the technology of making contacts. Almost all metals form silicides but the so called near noble metals palladium, platinum and nickel are of particular interest because they react with silicon at remarkably low temperatures. Evidence from UPS investigations in UHV indicates that interfacial reaction begins at room temperature [19]. The initial product when palladium reacts with silicon in the vicinity of 500K is Pd_2Si which is a hexagonal phase. The silicon atoms in the silicide have an interesting relationship to those in elemental silicon; their configuration in the (0002) planes of the silicide represents a sublattice of the sites in (111) in the silicon. There are two inequivalent (0002) planes in which 1/3 and 2/3 of the sites in the corresponding (111) planes are vacant. Pd_2Si grows epitaxially on $(111)_{Si}$ with a natural misfit at room temperature of 0.018. STM investigations have revealed microstructural details of the initial stages of the palladium-silicon reaction [20]. Fig. 6 is an STM image showing patches of Pd_2Si which have been formed by room temperature reaction between palladium and a (7x7) reconstructed $(111)_{Si}$ surface. The novel insight which comes from analysis of this image is that the hexagonal silicide has nucleated exclusively in the faulted i.e. hexagonally stacked half of the (7x7) supercell. This observation is one of the first to demonstrate the importance of a surface reconstruction in affecting the morphology of a growing film. By contrast the first step in the growth of niobium or molybdenum on reconstructed (100) or (111) surfaces is the reversion of the surface to the (1x1) state.

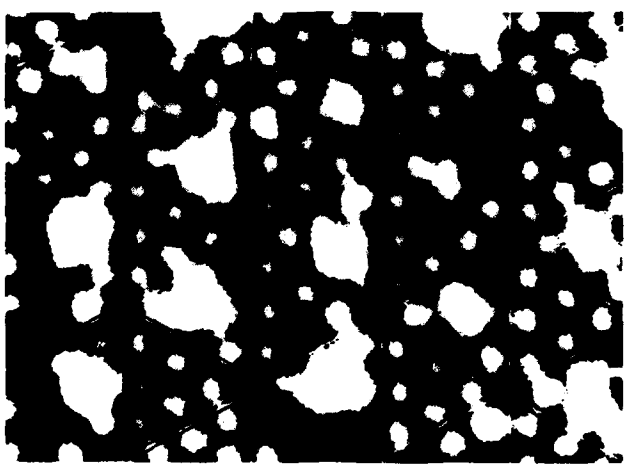


Figure 6: An STM image from ref. [20] of a 115Å x 95Å area of Si(111) - (7x7) with 0.25ML Pd. A grid is overlayed to indicate the substrate (7x7) lattice. The size of one (7x7) unit cell is highlighted in the lower left corner. The silicide is seen to nucleate exclusively on the faulted half of the unit cell (courtesy APS).

SCHOTTKY BARRIERS

One of the most important issues in semiconductor technology is the formation of metal-semiconductor interfaces and the height of the resulting Schottky barrier. Despite many years of study there is no clear understanding of the Fermi level pinning mechanisms involved, nor have the gap states which may be responsible for the pinning been clearly identified. Various models have been proposed including defect models with discrete gap states [21] which may be applicable to low metal coverages, and metal induced gap states (MIGS) [22] which may be important in the high coverage limit. The combination of the STM's high spatial resolution together with its spectroscopic capabilities make it potentially a very powerful technique for addressing these issues.

The most extensive spectroscopy has been performed on clean and oxygen covered GaAs(110) by Feenstra and Stroscio [6,23]. The clean surface has flat band conditions and the position of the Fermi level can be clearly seen in the I-V spectra. With the adsorption of oxygen, localized band bending can be seen around the oxygen in the STM images as well as the I-V spectra. This localized band bending was seen to decay away from the oxygen atoms, with a decay length of about 50Å due to screening from the conduction electrons. Although the presence of surface band bending is believed to result from surface states in the gap, none were seen in this study. In a more recent study of antimony on GaAs(110), Feenstra and Martensson [24] were able to spatially locate the states within the gap associated with the band bending resulting from antimony adsorption. At sub-monolayer coverages, the antimony forms islands. The gap states were found to be localized at the edges of these islands. The location of the gap states responsible for Fermi level pinning may help significantly in the understanding of the type of defects responsible for pinning and hence the formation of Schottky barriers. The STM results to date are limited to the low coverage limit, and it is not clear whether this work can necessarily be extended to the

understanding of the band bending involved with thick metal films.

BEEM (Ballistic Electron Emission Microscopy) is a recently developed variation on STM which permits the study of subsurface metal semiconductor interfaces [25]. In this technique tunneling occurs between the STM tip and the metal overlayer. However if the metal is only a few layers thick and the electrons have sufficient energy the injected ballistic electrons may propagate through the metal without scattering, into the semiconductor so probing the interface region. Although this technique has to be proven, it may provide a useful technique for studying the spatial variation of the interface electronic structure by measuring the ballistic current during scanning.

SURFACE MANIPULATION

Finally we turn to more directly applied uses for the STM. First can it be usefully employed as a routine analytical tool for imaging integrated device structures? STM images have successfully been made on microlithography patterns of 100nm in size [26]. However, accompanying SEM images of the same structure have better resolution! A significant drawback with the STM in this type of application is that the image is always a convolution of the tip and surface, and so if the surface is not flat to within a few atomic layers, the tip structure can become important and effectively reduce the resolution obtained. It may be that when microlithography produces features on the 10nm scale that the STM will be the only technique available, but its immediate use in this area does not seem promising. There is also considerable interest in the use of the STM as a tool for microfabrication or for writing information on a storage device. A few experiments on surface modification [27-30] have demonstrated the feasibility of writing features of a few nm in size on a surface and then imaging them. It may therefore be that the STM could provide a means of repairing circuits or fabricating particularly small scale features, but this may be somewhat restricted by the limited scan area that is practical, so making the location of a particular part of a device difficult. Similarly speed restrictions may limit the possibilities of practical STM based storage devices.

DISCUSSION

STM is clearly a potent addition to the existing techniques which can attain atomic resolution. Since the formation of such an image involves the tunneling of electrons to or from localized empty or filled states respectively proper interpretation requires a knowledge of the relation between the surface density of states and the atomic sites. For metals the density of states follows the ion cores but in semiconductors this need not be the case as the investigation of GaAs (110) demonstrates.

The STM has an advantage relative to high resolution transmission electron microscopy since there is no requirement that the sample has a simple stacking parallel to the electron beam. Of course this gain comes about because the STM is a surface technique. However the influence of surface relaxations needs to be evaluated when a lattice defect such as a dislocation or grain boundary is the subject of investigation. This same problem arises in the study of bulk defects by FIM; here there are further complications which arise from electrostatic forces comparable with the ideal strength of the material. Nevertheless the ability to characterize arbitrary defects without limitations being imposed by instrumental requirements such as periodicity in the electron beam direction in the HRTEM or the specimen in the form of a point for FIM, is a significant advance. It should also be remarked that, relative to FIM, the STM has the considerable advantages that the surface to be studied can be selected according to the priorities of the investigator and that all atom sites in that surface can be addressed. The addition of a capability corresponding to field evaporation would present a very desirable enhancement to the STM. A step towards the realization of this concept would be to combine STM and SIMS. It has already been demonstrated that the STM can distinguish atoms which are inequivalent in an electronic sense. These may, but need not be chemically distinct. So far the STM has been used to distinguish two or more known surface species but cannot as yet identify unknown species. The development of inelastic tunneling spectroscopy offers some potential in this direction.

In many aspects of the behavior of materials the path a system follows during a phase transition is as significant as the characterization of the final state. Consequently there is a strong motivation for in situ studies of the mechanisms and kinetics of solid state processes. Such investigations are an active area of research in transmission electron microscopy where crack nucleation [31], dislocation dynamics [32] and crystallization [33] are the objects of current attention. Clearly speed of data acquisition relative to the kinetics of the process concerned is a major consideration and any scanning technique is at a disadvantage relative to one where the image is parallel recorded. Even so surface diffusion has been observed by STM [34] and it may be anticipated that STM will in due course be applied successfully to measure surface diffusion coefficients at temperatures which are inaccessible to most other techniques.

CONCLUSION

The STM is already accepted as a powerful addition to the set of tools available to the surface scientist. Its especial strength lies in its ability to characterize surface structure and to some extent chemistry at the atomic level in an environment which is compatible both with complementary tools and a variety of processing options. It is anticipated that future developments may particularly emphasize studies of surface processes such as thin film growth and reactions.

ACKNOWLEDGMENTS

We would like to thank R. M. Feenstra, J. E. Griffith and U. K. Kohler for providing STM images.

REFERENCES

[1] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. 49, 57 (1982).
[2] M. D. Pashley, K. W. Haberern, and W. Friday, J. Vac. Sci. Technol. A6, 488 (1988).
[3] Sang-il Park and C. F. Quate, J. Appl. Phys. 62, 312 (1987).
[4] J. Tersoff and D. R. Hamann, Phys. Rev. Lett. 50, 1998 (1983).
[5] R. M. Feenstra, J. A. Stroscio, J. Tersoff, and A. P. Fein, Phys. Rev. Lett. 58, 1192 (1987).
[6] R. M. Feenstra and J. A. Stroscio, J. Vac. Sci. Technol. B5, 923 (1987).
[7] R. J. Hamers, R. M. Tromp, and J. E. Demuth, Phys. Rev. Lett. 56, 1972 (1986)
[8] J. A. Stroscio, R. M. Feenstra, D. M. Newns, and A. P. Fein, J. Vac. Sci. Technol. A6, 499 (1988).
[9] G. Binnig, H. Rohrer, Ch. Gerber, and E. Weibel, Phys. Rev. Lett. 50, 120 (1983)
[10] K. Takayanagi, Y. Tanishiro, M. Takahashi, and S. Takahashi, J. Vac. Sci. Technol. A3, 1502 (1985)
[11] R. M. Tromp, R. J. Hamers, and J. E. Demuth, Phys. Rev. B34, 1388 (1986).
[12] M. D. Pashley, K. W. Haberern, W. Friday, J. M. Woodall, and P. D. Kirchner, Phys. Rev. Lett. 60, 2176 (1988).
[13] M. D. Pashley, K. W. Haberern, and J. M. Woodall, J. Vac. Sci. Technol. B6 (1988) (to be published).
[14] D. J. Chadi, J. Vac. Sci. Technol. A5, 834 (1987).
[15] P. E. Wierenga, J. A. Kubby, and J. E. Griffith, Phys. Rev. Lett. 59, 2169 (1987).
[16] J. E. Griffith, J. A. Kubby, P. E. Wierenga, R. S. Becker, and J. S. Vickers, J. Vac. Sci. Technol. A6, 493 (1988).
[17] J. E. Griffith, J. A. Kubby, P. E. Wierenga, and G. P. Kochanski, (to be published).
[18] R. S. Becker, B. S. Swartzentruber, J. S. Vickers, M. S. Hybertsen, and S. G. Loiue, Phys. Rev. Lett. 60, 116 (1988).
[19] G. W. Rubloff, P. S. Ho, J. F. Freeouf, and J. E. Lewis, Phys. Rev. B23, 4183 (1981).
[20] U. K. Kohler, J. E. Demuth, and R. J. Hamers, Phys. Rev. Lett. 60, 2499 (1988).
[21] W. E. Spicer, I. Lindau, P. Skeath, C. Y. Su, and P. Chye, Phys. Rev. Lett. 44, 420 (1980).
[22] J. Tersoff, Phys. Rev. Lett. 52, 465, (1984).
[23] J. A. Stroscio, R. M. Feenstra, and A. P. Fein, Phys. Rev. Lett. 58, 1668 (1987).
[24] R. M. Feenstra and P. Martensson, (to be published).
[25] W. J. Kaiser and L. D. Bell, Phys. Rev. Lett. 60, 1406 (1988).
[26] S. Okayama, M. Komuro, W. Mizutani, H. Tokumoto, M. Okano, K. Shimizu, Y. Kobayashi, F. Matsumoto, S. Wakiyama, M. Shigeno, F, Sakai, S. Fujiwara, O. Kitamura, M. Ono, and K. Kajimura, J. Vac. Sci. Technol. A6, 440 (1988).
[27] U. Staufer, R. Wiesendanger, L. Eng, L. Rosenthaler, H. R. Hidber, H.-J. Guntherodt, and N. Garcia, Appl. Phys. Lett. 51, 244 (1987).
[28] R. M. Silver, E. E. Ehrichs, and A. L. de Lozanne, Appl. Phys. Lett. 51, 247 (1987).
[29] R. S. Becker, J. A. Golovchenko, and B. S. Swartzentruber, Nature 325, 419 (1987).
[30] J. S. Foster, J. E. Frommer, and P. C. Arnett, Nature 331, 324 (1988).
[31] Y. H. Chiao and D. R. Clarke, Acta Metall. (to be published).
[32] R. Hull and J. C. Bean, Ultramicroscopy (to be published).
[33] D. A. Smith, K. N. Tu, and B. Z. Weiss, Ultramicroscopy 23, 405 (1987).
[34] R. C. Jaklevic and L. Elie, Phys. Rev. Lett. 60, 120 (1988).